ELUCIDATING HOW GLOBAL CLIMATE CHANGE FACTORS AFFECT SOIL MICROBIAL CARBON CYCLING PROCESSES: FROM TROPICAL FORESTS TO THE ALASKAN TUNDRA

A Dissertation
Presented to
The Academic Faculty

by

Eric Robert Johnston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
December 2018

ELUCIDATING HOW GLOBAL CLIMATE CHANGE FACTORS AFFECT SOIL MICROBIAL CARBON CYCLING PROCESSES: FROM TROPICAL FORESTS TO THE ALASKAN TUNDRA

Approved by:

Dr. Konstantinos T. Konstantinidis,
Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Joel Kostka
School of Biological Sciences
Georgia Institute of Technology

Dr. Spyros G. Pavlostathis
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. James M. Tiedje
NSF Center for Microbial Ecology
Michigan State University

Dr. Joe Brown
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: August 09, 2018

ACKNOWLEDGEMENTS

First, I would like to express my sincere graditude to my advisor, Dr. Kostas Konstantinidis, for mentoring me throughout my graduate studies at The Georgia Institute of Technology. I will be forever grateful for the opportunity to stand beside him at the frontier of science and engineering, and for how he has shaped me into a hard worker and a more critical thinker. I am also gradeful to the members of Kostas' lab group for their support and assistance. In particular, I would like to thank Dr. Luis Miguel Rodriguez-R for the numerous cutting-edge analysis tools he has made available to his peers and Minjae Kim for his friendship and assistance during my thesis work. I would also like to thank Dr. Ximei Zhang, whose collaboration on several research projects has been critical to my continued success and to my dear friend Joshua Bierdz, whose converstations about life and philosophy has helped fuel and refine my own personal objectivity. Finally and most of all, I would like to thank my partner, Brittany Anaya, whose love and support was essential throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Soil systems are residence to perhaps the most diverse biological communities on Earth, composed of thousands of distinct microbial populations, each typically representing a small fraction of the total community. These complex biological systems play an important role in numerous ecosystem functions that are vital to humans and other organisms, such as nutrient cycling, organic matter stabilization, food and fiber production, water quality maintenance, and waste remediation. Anthropogenic induced changes to the environment, such those resulting from climate change, global nitrogen deposition, or land conversion for agroecosystem use, threaten to impair or alter these important functions. How soil communities respond or adapt to such globally prevalent disturbances remains poorly understood. Hence, there is an imminent need to evaluate soil functioning and resilience under different types of environmental change in order to better assess their long-term impacts. The results of such efforts could inform policy and lead to improved land management practices aimed at mitigating the negative consequences of environmental change.

In particular, it remains unclear how climate warming will change long-term global soil carbon (C) storage. Soils harbor a massive reservoir of C that is several times greater than the amount present in the atmosphere, mostly in the form of organic C (OC) materials susceptible to microbial degradation. There is growing concern that climate warming will stimulate microbial activities involved in C release, which has motivated research efforts into studying the effects of climate change on soil systems. Due to their importance in global C cycling and anticipated responsiveness to rising temperatures, efforts have also been made

to represent microbial soil C turnover processes into climate models. But because our current understanding and representation of these processes are inadequate, model projections are highly variable and considered to be of low confidence. These challenges highlight how an improved understanding of soil systems and their microbial inhabitants is necessary for improving predictions of future climate change. **Chapter 1** covers many of the outstanding challenges regarding modeling of soil C dynamics in greater detail.

Our ability to study and predict soil microbial responses is hindered by the enormous diversity and unculturability of soil microorganisms. However, a new generation of advanced 'omics techniques (e.g., genomics, transcriptomics, proteomics) provides improved methods for investigating how taxonomic, genetic, and functional attributes of soil microbiota respond to global change factors and how these responses relate to measured processes rates (e.g., CO_2 and CH_4 emissions). Making use of these techniques, we have investigated soil microbial communities from several ecosystems and their responses to relevant climate change factors over time under both field and laboratory incubation conditions. These investigations have included, for instance, soil communities inhabiting tropical rainforests where plant CO_2 uptake is high, but where C cycling is also often constrained by phosphorus (P) availability (**Chapter 2**). Soil microorganisms are essential to inorganic P mobilization and organic matter turnover, but C-P-coupled biogeochemistry in tropical soils remains an underexplored area of research. Our investigation of these systems involved quantifying respiration between soils with varying degrees of P limitation and elucidated major physiological mechanisms underlying P-mediated OC turnover.

As part of a 5 year climate change experiment conducted in a semi-arid, temperature-limited steppe, we revealed how warming stimulates microbial

mechanisms involved in C release, but also how increased precipitation expected in this ecosystem could counteract C loss through increased plant productivity and OC inputs to the soil (**Chapter 3**). Similarly, we have investigated microbial communities of Alaskan tundra soils, where C stocks are considered to be particularly vulnerable to rising temperatures, which have undergone experimental warming *in situ* (~1-2 °C above ambient) for several years. Deep sequencing of these communities allowed for a unique opportunity to assess the extent of their natural diversity, and for comparison with soil communities inhabiting temperate regions (**Chapter 4**). These datasets also facilitated the recovery of several genomes representing novel taxa, many of which were found to be widespread throughout the surrounding ecosystem. Through time series monitoring of these soil communities, we identified how future warming can be expected to stimulate their functions involved in CO_2 and CH_4 emissions (**Chapter 5**). Our studies demonstrated not only the rapid responsiveness of tundra soil communities to rising temperatures, but also how contrasting modes of response can be expected at different depth profiles, such as recently thawed permafrost vs. pre-existing active layer communities. Such contrasting responses also accentuate the potential shortfalls of estimating ecosystem-wide responses with oversimplified expressions commonly used by climate models.

These investigations contribute to an improved understanding of the diversity and functionality of terrestrial soil microbial and their responses to changing temperatures, precipitation patterns, and nutrient conditions. Such genomic indicators and physiological responses reported in these investigations have helped lay the necessary groundwork for further experimentation, in addition to broadening our understanding of soil microbial communities indigenous to each of these highly unique ecosystems.

CHAPTER 1: TOWARDS AN IMPROVED REPRESENTATION OF SOIL MICROBIAL CARBON CYCLING PROCESSES IN ECOSYSTEM MODELS

Eric R Johnston & Konstantinos T Konstantinidis

1.1. Introduction to climate change, the global carbon cycle, and soils

The acceleration of global climate change, driven primarily by a ~50% increase in atmospheric carbon dioxide (CO_2) levels from the preindustrial era to modern day, remains one of the greatest scientific and political concerns of the 21st century (IPCC 2014). Current CO_2 concentrations surpass those from any point in the past 800,000 years, which remained at or below 300 ppm (Lüthi et al., 2008), and may even reflect the highest point in 2.4 - 3.3 million years (Tripati et al., 2009). Anthropogenic fossil fuel consumption and land use change are the main drivers of this increase; with cumulative C release estimated as 570 ± 70 Pg C (Joos and Spahni 2008; Le Quéré et al., 2014). However, this amount is much greater than what is reflected by increased atmospheric CO_2. The discrepancy is primarily attributed to two substantial negative feedbacks mitigating atmospheric CO_2 influx: enhanced CO_2 fertilization of terrestrial plants and uptake by marine systems. Each of these mechanisms has offset atmospheric CO_2 increase by 25-30%, and together mitigates approximately half of all C emissions (Le Quéré et al., 2014).

Terrestrial primary producers use approximately 46% of this fixed organic carbon (OC) for their own consumption, while the remainder enters the soil organic C (SOC) pool as detritus. There, it is susceptible to decomposition by heterotrophic microorganisms. However, SOC can persist as undegraded plant carbon, microbial biomass or necromass, metabolic by-products, and a variety of other compounds. As a result, soils harbor a substantial amount of C, two-thirds of which is the form of undegraded organic matter (Ciais et al., 2013). It has been estimated that the amount of C stored in arctic tundra soils alone, which make up just ~16% of Earth's land surface, is greater than the amount of aboveground plant and atmospheric C combined (~1,672 Pg C) (Tarnocai et al., 2009). Microbial respiration of SOC results in an annual C flux of 60 - 75 Pg C, which is over an order of magnitude larger than the current ~4.3 Pg annual increase in atmospheric C (Ciais et al., 2013; Le Quéré et al., 2014). Due to the delay between elevated greenhouse gas concentrations and expected maximum warming response, biological responses to elevated temperatures and CO_2 can proceed for centuries, even if C emissions were to abruptly cease (Zickfeld and Herrington 2015). Hence, changes to soil C flux could have large implications for the global carbon cycle and future climate change, particularly if these changes are maintained for long durations.

The possibility that climate warming will stimulate microbial decomposition of SOC, leading to enhanced CO_2 and CH_4 release, has motivated research efforts into studying the direct effects of elevated temperatures on soil systems (Zhou et al., 2011). Thus, despite greater sustained production of terrestrial vegetation and increased inputs to the soil, the threat of enhanced microbial decomposition may reduce the capacity of terrestrial systems to remain a C sink or eventually become a source and accelerate climate change (i.e., positive

feedback) (Arora et al., 2013; Friedlingstein et al., 2014). Due to a disproportionately large C reservoir and comparatively high susceptibility to rising temperatures, tundra systems are at particularly high risk of becoming a C source (Schuur et al., 2008). However, a change in terrestrial soil carbon storage ultimately depends on the balance between photosynthesis, respiration from soil heterotrophs, and C soil stabilization over time. An improved understanding of the dominant factors regulating the fate of soil C is essential for more accurate predictions of future climate change to emerge.

1.2. Inclusion of soil C processes in climate models

Due to the large interactions between major C pools (e.g., atmosphere, terrestrial plants, soils, etc.), many of the most sophisticated and commonly used climate models, such as the Earth Systems Models (ESMs), incorporate the movement of C between these systems (Ciais et al., 2013). Our ability to predict future climate change relies heavily on the adequate portrayal of various feedback mechanisms that can amplify or weaken its effects. Soil C feedbacks have been regarded as one of the greatest areas of uncertainty in climate modeling, comparable perhaps only to the uncertainties surrounding cloud feedbacks (Gregory et al., 2009; Friedlingstein et al., 2014). This is attributed not only to the large reservoir of C stored in soils, but also to an inadequate representation of soil C processes by climate models. While the past few decades have resulted in a much-improved portrayal of the physical mechanisms involved in climate change by ESMs (Knutti and Sedláček 2013; Flato et al., 2013; Stocker et al., 2013), the representation of soil C turnover has essentially remained unchanged (Jenkinson et al., 1991; Todd-Brown et al., 2014). This lack of progress is despite recent technological breakthroughs during the same period

that have propelled our ability to investigate and better understand soil microbial processes involved in C cycling (e.g., high-throughput DNA sequencing and other emerging molecular-based approaches). Microbial activities mediating the release of C are commonly implied as a first-order response curve, where the time taken to decompose a fraction of SOC increases with elevated temperatures (Todd-Brown et al., 2013; Todd-Brown et al., 2014; Sierra et al., 2012; Xia et al., 2013). Some models include a more complex representation of soil C dynamics than others, involving two or more C pools with contrasting decomposability, moisture, texture, and other soil properties (Luo et al., 2016). When attempts are made to include these simple portrayals of soil microbial response into ESMs, estimates of warming-induced changes occurring between now and 2100 range from a slight increase in soil C to one-third of the total stock emitted (Arora et al., 2013; Friedlingstein et al., 2014; Cox et al., 2013). There is also low confidence that the range or aggregate mean of these projections hold any significance, because the representations of soil processes common amongst model implementations are inadequate (Bradford et al., 2016). Thus, there is a need to better understand the mechanisms governing SOC turnover and persistence, so they can be identified and scaled for use in ESMs. Such advancements are necessary for reducing uncertainty in future SOC stock projections.

1.3. Persisting uncertainties regarding soil temperature sensitivity

One major implicit representation of soil microorganisms in climate models is through Q_{10}, a parameter describing the increase in soil respiration coinciding with a 10 °C increase in temperature, which has been thought to scale up well for capturing macro-level patterns of response (Davidson and Janssens 2006; Lloyd and Taylor 1994). However, when using field data-informed Q_{10} across a broad

range of climate model estimates as part of the Couples Model Intercomparison Project Phase 5 (CMIP5), projected changes to soil C storage by 2100 ranged from -70 to 250 Pg (Todd-Brown et al., 2014). Further, while this portrayal of soil C temperature-sensitivity seems to fit the behavior of soils well over short durations, a recent synthesis of various global warming experiments implies that implementation of Q_{10} is potentially inadequate (Carey et al., 2016). This is because of a broad decline in temperature sensitivity around 25 °C observed for most soils, after which respiration rates seemingly reverse. A common decline in soil moisture with elevated temperatures accounted for some of this reduction, but a similar decline in response was nonetheless observed even under the wettest conditions. Use of Q_{10} infers an exponential increase in respiration with rising temperature, and the synthesis by Carey et al., 2016 demonstrated how this relationship does not hold across the full temperature range. Thus, ESMs implementing Q_{10} could be overestimating respiration rates.

Another major finding by Carey et al., 2016 was that several years of experimental warming of various soils did not result in altered temperature sensitivities. This could lead one to conclude that microbial communities may not acclimate nor become optimized to higher temperatures. However, it was also demonstrated that a vastly different threshold for reduced respiration was observed for desert ecosystems (at 55 °C vs. 25 °C for almost all other environments). The authors attributed this conflicting result to two possible explanations: either plant and microbial species inhabiting deserts are physiologically adapted to higher temperatures, or because of the comparatively greater role of UV in photodegradation of organic compounds in these systems (Balser and Wixon 2009; Austin and Vivanco 2006). The former explanation conflicts not only with the notion of 25 °C representing a near-universal

temperature optimum, but also with the unchanging temperature sensitivities of soils subjected to several years of experimental warming. Not represented in this synthesis were experimentally warmed tropical soils, for which no datasets existed at the time of writing. Tropical rainforests at the equator support some of the greatest rates of primary production and have an average annual temperature of up to 30 °C, beyond the abovementioned suggested thermal optimum. A universal decline in respiration rates past 25 °C contradicts how SOC is more rapidly consumed in tropical forest soils compared to those in temperate forests, despite having mean annual temperatures at or exceeding 25 °C and extremely high productivity rates (Wang et al., 2018). The notion that optimal temperature responses are disconnected from microbial physiological adaptations also conflicts with a multitude of studies demonstrating how enzymes of microbiota reflect local temperature constraints, such as those encoded by psychrophilic taxa adapted to more frigid conditions (D'Amico et al., 2006).

The greatest experimentally-induced average temperature increase of any biome used in the synthesis was 1.9 °C (for grasslands). Topsoils can experience a broad range of seasonal temperature variation, and thus, could possess microorganisms uniquely adapted to different points across this range. The magnitude of temperature increase used in most of these warming experiments likely does not place soil microorganisms far outside of their traditional temperature range except during seasonal extremes. With this in mind, the majority tundra habitats have already experienced a 2 - 4.2 °C temperature increase during the past half-century alone (2008-2017 vs. 1950-1980 base period) (Hansen et al., 2010; GISTEMP Team, 2018), and it is expected that these temperatures will further increase. Accelerated warming over the past fed decades in northern high-latitude regions has resulted in rapid permafrost

degradation (Schuur et al., 2008; Schuur et al., 2009). It has been projected that permafrost will further recede by 30-70% by the end of the 21st century (Lawrence et al., 2012). Microbial decomposition of previously preserved permafrost C represents a potential feedback that could be particularly potent, because thawing induced by warming can also make permafrost a substantial methane source (Liebner and Wagner 2007; Schuur et al., 2008; Schuur et al., 2009). Methane concentrations have been increasing over the past decade, and this is likely due to, at least in part, to warming-induced release of methane from tundra soils. Also in contrast to the suggestions of Carey et al., 2016 that microbial communities to not adapt to increased temperatures, a previous incubation involving high latitude soils found that respiration temperature sensitiveity increased by a factor of 1.4 following a 90 day adjustment period (Karhu et al., 2014). While it can be assumed that incubation conditions do not reflect community adaptations that would occur in natural systems, partly because such conditions don't allow for the migration of new members, an adjustment of the existing community was observed nonetheless. **Hence, whether adaptation plays a negligible role in long-term community temperature response should be evaluated further. This should involve experiments using greater temperature increases over greater durations, with a particular focus on temperature-constrained habitats**.

Tropical systems remain poorly represented by long-term field warming experiments, perhaps because of an assumed absence of temperature constraints. However, such experiments, particularly in the warmest regions at equatorial rainforests, could provide important insight on how C cycling might change at very high temperatures, and specifically, whether a global ~25 °C optimum for respiration exists and can be exceeded. Inclusion of tropical soils

could also facilitate cross-biome comparisons to elucidate broad ecological patterns between soils subject to different temperatures or rates of aboveground production. The above discrepancies highlight areas of outstanding uncertainty regarding long-term temperature responses of soils.

While making inferences about community adaptations to elevated temperatures, another outstanding problem is the ability to discern between short and long-term responses induced by warming. In most warming experiments, soils are typically subjected to a prolonged period of elevated temperatures, usually years, after which experimentally warmed or unwarmed soils are collected at the same time for evaluation and compared. As soil communities are undoubtedly responsive to seasonal fluctuations in the environment, it can be expected that there exist natural fluctuations in the composition and functioning of these communities. Thus, it is possible that short-term responses in community structure as a result of warming occur, and that they may also be overlap in structural and physiological responses between warmed and unwarmed soils when temperatures are comparable (i.e., at different points in the year). Through an improved evaluation of natural intra-annual responses, these seasonal fluctuations could be distinguished from the long-term effects of warming, allowing for better interpretations regarding the effects of warming to emerge (i.e., on microbial community structure, and how this relates to observed ecological patterns and physiological response).

1.4. Improving soil C predictions for when bottom-up controls are weaker

The importance of bottom-up controls over SOC (e.g., environmental conditions, such as temperature) is well established and remains a common

focus of research and modeling efforts. For instance, low temperatures constrain the turnover of plant detritus, causing these materials to accumulate and persist in frigid tundra soils, as has been validated empirically (Hofle et al., 2013). For habitats with fewer bottom-up controls, the dominant mechanism of SOC persistence has traditionally been summarized as preferential consumption of labile plant compounds and an accumulation of comparatively more recalcitrant plant compounds (Berg and McClaughery 2010; Waksman 1936). In either case, the common portrayal of SOC is that its persistence depends on environmental conditions or inherent chemical properties that restrict its decomposition. The contribution of microbial processed and biosynthesized OC has always been recognized, but only recently have these materials been suggested to represent a dominant, or even, the dominant mechanism of OC formation and persistence. There is accumulating evidence that SOC is dominated by microbially-processed materials in systems without strong bottom-up controls (Paul 2016). It was recently shown that C- and microbe-free soils supplied with simple carbon substrates and inoculated with microorganisms from natural soil eventually displayed a broad chemical profile, including aromatics, protein, lipids, chitin, phenolics, and polysaccharides, roughly reflecting the broad chemical composition of field soil after a period of just 18 months (Kallenbach et al., 2016). It has further been suggested that as much as 80% of SOC may actually reflect microbial biomass, necromass, and metabolic by-products of microbial metabolisms in certain soils (Liang and Balser 2011).

Studies examining the effects of accelerated soil respiration (due to warming or other factors) have demonstrated a common pattern, where enhanced SOC turnover does not confer to a linear increase in accumulated microbial products. These observations have been attributed to a few possible

underlying mechanisms that are somewhat overlapping in their implications. For instance, it has been suggested that warming can decrease microbial growth efficiency (MGE) by increasing the cost of maintaining existing biomass at higher temperatures (Manzoni et al., 2012; Sinsabaugh et al., 2013; Dijkstra et al., 2011). Enhanced decomposition has also been associated with greater microbial turnover (Sinsabaugh et al., 2013). Use of isotope labeled substrates allowing for MGE and rates of microbial turnover to be evaluated separately implied that enhanced microbial turnover was the prevailing response to elevated temperatures that facilitated enhanced SOC turnover (Hagerty et al., 2014). A positive relationship between microbial biomass C and accelerated microbial turnover with increased respiration has also been reported (Kaiser et al., 2014). Another way to evaluate the activity-biomass relationship is through the microbial metabolic quotient (MMQ), which is defined as microbial respiration per unit biomass (Anderson and Domsh 1993). A recent study demonstrated how increasing MMQ at elevated temperatures is a globally widespread phenomenon across different soils and was also correlated with soil texture (Xu et al., 2017).

As environmental conditions become less constraining and plant inputs increase (as is typically the case with warmer temperatures and elevated CO_2), causing the composition of SOC to shift from undegraded plant to microbial materials, the question becomes, “what controls the persistence of microbially processed SOC?”. The mechanisms that limit SOC accumulation as assimilated microbial biomass and by-products, in particular, remains an area of uncertainty and is not well reflected by current ESM implementations. This is an important topic to address, because rising temperatures and enhanced plant productivity is expected for the majority of habitats. One commonly overlooked factor potentially governing these responses is the turnover of microbial compounds undoubtedly

facilitated by other organisms responsible for their decomposition, such as micropredators and scavengers. Such life strategies could be more resource intensive than in other environments, such as aquatic systems, due to the need to navigate throughout a complex soil matrix in search of microbial prey and necromass. Thus, the involvement of these taxa in C turnover is then likely dependent on the density of other cells, such as those of primary consumers. Studies of experimentally warmed soils found inconsistent responses in the abundances of predator and grazer organisms as an effect of warming (Blankinship et al., 2011). However, representation of predator abundance was limited to soil fauna, which neglects the role of predator-prey relationships that might exist between different bacterial groups.

Although soil bacterial predator-prey interactions remain largely underexplored, the complex strategies required might be well reflected in the few bacterial micropredators that have been intensively studied, one being *Myxococcus xanthus*. Myxobacteria are ubiquitous in soils, having been found on every continent, excluding Antarctica, and are estimated to make up 0.4% to 4.5% of all soil Bacteria based on a collection of 182 soil samples (Zhou et al., 2014). Myxobacteria are obligate social predators known to hunt in ‘wolf-packs’. During periods of starvation, myxobacteria populations on the order of 100,000 cells form a complex fruiting structure while waiting for prey to accumulate in the surrounding environment (O'Connor and Zusman 1989). When molecular signals released by potential prey accumulate to sufficient levels, the bundle of myxobacteria cells become active and swarm outwards in all directions. While doing so, they secrete a wide array of antibiotics and digestive enzymes that can allow for the complete digestion of both gram-negative and gram-positive prey (Berleman et al., 2014). These social predators even exhibit phenotypic

differentiation, with individual cells conferring different functions as a suggested division of labor (van Vliet and Ackermann 2015). After digestion of surrounding prey is completed and the cells begin to experience starvation, they secrete chemicals signaling for the swarm to congregate and reform the fruiting structure. There, they become dormant again and wait for local prey to replenish before repeating this process (Dahl et al., 2011). This coordination between cells in forming the fruiting body is an energy intensive process, causing up to 90% of cells to undergo autolysis to provide sustenance for other cells between cycles of dormancy and predation. Another well-studied bacterial micropredator, *Bdellovibrio bacteriovorus*, is ubiquitous in bulk and rhizosphere soils (Jurkevitch et al., 2000), and exhibits a predatory pattern very different to that of myxobacteria (Sockett 2009). *Bdellovibrio* cells are comparatively smaller, allowing them to grow and replicate within the periplasm of larger prey, using their cytoplasmic contents and cell structures as an obligatory source of energy and nutrients (Hespell et al., 1973). **Due the high complexity of soil microbial communities and the existence of small and large predators possessing contrasting strategies of predation, it is highly possible that soils feature a cyclic (non-unidirectional) food web that can serve as a funnel for energetic resources, driving C loss from the system under conditions of biomass accumulation.**

If environmental controls do not limit heterotrophic consumption of plant-derived OC inputs, perhaps the SOC pool itself becomes self-limiting as a result of density-dependent trophic interactions. Such an overlooked dynamic could have big implications for soil C cycling, serving as an inherent control over soil C accumulation. This would not necessarily imply that such interactions constrain the activities of primary consumers; instead, it could cause the opposite.

Competition for physical space or limiting nutrients resulting from overcrowding of primary consumers could impair the further degradation of plant materials. Such a result was observed in experimentally warmed Harvard Forest and Duke Forest soils, where warming-induced increases to soil respiration were strongly facilitated by the presence of soil predators (but like other studies, this assessment was limited to invertebrate organisms and did not include bacterial predators) (Pelini et al., 2015). It is highly likely that previous observations of increased plant inputs, temperatures, or otherwise elevated soil respiration resulting in lower MGE and increased microbial turnover could be more directly attributed to an expanding food web. While there is some overlap in these interpretations and potential similarity in how they could be modeled to predict soil C dynamics, incomplete theory and indirect evaluation does not help much in reducing imprecision and uncertainty.

The representation of soil C loss stemming from trophic level energy transfer may seem too burdensome for implementation in ESMs. However, what if the general trends underlying these ecological mechanisms could be captured in a very simple way, and made use of information already collected for estimating the bottom-up controls on plant OC turnover? This could involve a model framework that includes microbial cell density (i.e., microbial biomass carbon), as well as the major factors affecting micropredation efficiency (e.g., texture, moisture, nutrient limitation). With inclusion of these variables, perhaps the presence of some broadly-scalable macro-level pattern would be observed. Implementation of this component could account for uncertainties that arise for environments with fewer bottom-up limitations and increased decomposition of plant matter does not result in a linear increase to measured microbial materials.

Efforts should be undertaken to see if such patterns prevail and to determine how applicable they are between soils. For this, soils could be amended with isotope labeled organic substrates of varying input designed to both underwhelm and overwhelm the soil community and evaluated with continuous measurements of soil respiration and microbial biomass. With a time series evaluation of communities enriched with the isotopically labeled substrates (e.g., using DNA sequencing approaches), one could evaluate the flow of carbon through a food web. Microbiota enriched in later time points after the substrate has been depleted and in the overwhelmed system are likely to represent soil micropredators and scavengers. The abundances and/or activities of these micropredators could then be related to measured microbial turnover rates or MGE. If such a relationship between microbial turnover or MGE and enrichment of soil predators is established, soils with different properties limiting or facilitating predation (i.e., texture, moisture) could then be used to assess if and how the underlying mechanisms could be incorporated into a common model framework. Additionally, use of fluorescent in situ hybridization (FISH) probes could be used to validate predator-prey predictions inferred through co-occurrences patterns or isotope label enrichment. However, application of FISH to study soil microbiota has been challenging, as weak signal intensities and high levels of autofluorescence emitted by soil particles and plant tissue make it difficult to discriminate between targeted microbial cells and other molecules in the soil environment (Briones et al., 2002; Bloemberg 2007; Buddrus-Schiemann et al., 2010). Further advancements in methods for use of FISH to study soil microbiota would better facilitate the widespread adoption of these techniques.

One important consideration in the effects of predator-prey relationships on soil C storage is the role of soil community diversity. While the well-studied

Myxococcus xanthus appears to have low prey specificity, this may not be true for all or even most micropredators. The effect of highly specific predator-prey associations could be reduced by higher levels of community diversity, where the cell density of unique prey populations would typically be lower, leading to an overall constraint on predation and increase in soil C storage. Alternatively, a greater diversity of organisms could select for predators with lower prey specificity, and thus, lower overall metabolic efficiency, causing greater C loss for each trophic-level transfer of materials and energy.

Capturing SOC dynamics representing the most essential processes and interactions that are globally-scalable in order to remain manageable for parameterization in ESMs remains an ongoing challenge. However, current models do not adequately depict SOC turnover rates, particularly when environmental constraints are reduced and organic inputs are high, and thus, their projections probably do not match what can be expected from long-term outcomes. Others have found associations between enhanced microbial activity and CUE or microbial turnover rates, or even interpreted these observations as differences between copiotrophic and oligotrophic life strategies. **We hypothesize that density-driven trophic level energy transfer may represent a major constraint on SOC that becomes increasingly relevant when bottom-up controls are alleviated and OC inputs are high, and can explain these common, overlapping observations of previous studies.**

1.5. Adequate representation and inference of soil microbial materials

Recent advances in the recovery and characterization of microbial materials (DNA, proteins, transcripts, etc.) have revolutionized the study of

environmental systems. For instance, in the case of soil C storage, they can be used to infer the lifestyles and responses of organisms acting as the primary mediators soil C release. However, despite the vast number of studies now making use of these ‘omics techniques to address a variety of ecological questions, the degree to which recovered materials adequately reflect the original environmental sample remains speculative (Nesme et al., 2016). Despite hundreds of published articles evaluating community structure from recovered DNA, the determined structures have not been validated. This is also problematic because no benchmark allows for the complete assessment of biases that arise between differenct workflows. For instance, a direct comparison of community composition from the same soil collection using several different DNA extraction techniques found that the number of bacterial genera recovered differed from ~50 to 200 (Delmont et al., 2010). By pooling together DNA isolated from different methods, there was a further 80% increase in total genera richness. It has also been shown that the kit used for DNA library preparation (a necessary step prior to sequencing) can result in measured abundances of community members that differ by a factor of three (based on a mock community of 20 strains with cells in equal abundances) (Jones et al., 2015). As a separate confounding factor in interpreting the relevancy of recovered soil microbial materials, a recent study has found that up to 40% of soil DNA corresponds to extracellular material that has persisted in the soil habitat for years and can inflate bacterial diversity estimates by up to 50% (Carini et al., 2017). Due to the more recent nature and methodological difficulties involving transcript and protein assessments (i.e., transcriptomics and proteomics), it remains less clear how reflective these measurements are of the activities and conditions they are meant to evaluate. Biases and inefficiencies resulting from different methodological

workflows are even more problematic for efforts aimed at synthesizing information across studies, where differences between methodologies undoubtedly make datasets less comparable and perhaps intractable.

Rapidly declining costs of sequencing and more recently developed long-read technologies are increasing the resolution of DNA-based communities, and may eventually reach the point of genome-resolved metagenomes. Yet, such advancements in our ability to recover genes and genomes directly from environmental samples are beginning to far outpace advancements in our ability to characterize what has been recovered. Functional characterization of recovered sequences still relies on homology searches against references with verified functions. However, very few soil microorganisms have been cultivated, so many of their functions remain unknown. For this, improvements in microbial isolation techniques and innovative methods to explicitly assign functions to genomic elements and individual taxa remain important. But for the study of microorganisms that resist isolation for whatever reasons, such as those that intimately rely on the co-existence of other taxa, functional assignment may require highly advanced techniques to link functions to genes and taxa *in-situ*. The latter could be achieved, for instance, by integrating numerous highly resolved data types (RNA, DNA, proteins, chemical constituents, etc.) with advanced mathematical and computational techniques in a way that explicitly links the different data together with high confidence. Experimental deconvolution of complex communities (e.g., via bottlenecking diversity) may help facilitate these advancements. Nonetheless, many studies attempting to further our understanding of soil C responses to warming make use of these emerging technological solutions. Limitations in our ability to infer functioning and draw associations between data points are constrained by limitations stemming from

the technologies themselves, requiring further advancement in the methodological approaches as well as theory.

1.6. Summary and future recommendations

While the effect of heightened temperatures and carbon dioxide levels on other biological and physical systems has become increasingly resolved, our understanding of soil microorganisms and their roles in terrestrial C storage is not yet sufficient to adequately represent their contributions and responses to climate change. As such, several efforts have focused on identifying universal patterns that can be useful for implementation in ESMs. One outstanding problem remaining is the disentanglement of direct and indirect effect of warming on soil microbiota as well as their short vs. long-term responses. Certain vulnerable ecosystems, such as tundra habitats, should continue to be studied intensely, as the potential for C release from these systems disproportionately large. In particular, incorporation of soil C responses into ESMs should prioritize adequate representation of thawing permafrost systems. While models continue to focus on the bottom-up constraints of SOC turnover, there remains a need to better understand what controls the fate of SOC under long-term conditions where bottom-up controls are comparatively minor. Emerging research is showing that previous ideas regarding SOC formation and persistence may be inadequate to address this topic, meaning new ideas are needed to portray and summarize the important features involved. Our ability to predict the long-term fate of soil C with climate models involves synthesizing emerging theories in microbial ecology, representing the most pertinent microbial controls over soil organic matter turnover, and exploring methods that could be used to validate and parameterize these relationships.

1.7. References

Anderson T, Domsch K (1993) The metabolic quotient for CO2 (qCO2) as a specific activity parameter to assess the effects of environmental conditions, such as ph, on the microbial biomass of forest soils. Soil Biology and Biochemistry 25:393–395

Arora VK, Boer GJ, Friedlingstein P, et al (2013) Carbon–Concentration and Carbon–Climate Feedbacks in CMIP5 Earth System Models. Journal of Climate 26:5289–5314

Austin AT, Vivanco L (2006) Plant litter decomposition in a semi-arid ecosystem controlled by photodegradation. Nature 442:555–558

Balser TC, Wixon DL (2009) Investigating biological control over soil carbon temperature sensitivity. Global Change Biology 15:2935–2949

Berg B, McClaugherty C (2010) Plant Litter: Decomposition, Humus Formation, Carbon Sequestration (Springer, 2010).

Berleman JE, Allen S, Danielewicz MA, et al (2014) The lethal cargo of Myxococcus xanthus outer membrane vesicles. Frontiers in Microbiology 5

Blankinship JC, Niklaus PA, Hungate BA (2011) A meta-analysis of responses of soil biota to global change. Oecologia 165:553–565

Bloemberg GV (2007) Microscopic analysis of plant–bacterium interactions using auto fluorescent proteins. European Journal of Plant Pathology 119:301–309

Bradford MA, Wieder WR, Bonan GB, et al (2016) Managing uncertainty in soil carbon feedbacks to climate change. Nature Climate Change 6:751–758

Briones AM, Okabe S, Umemiya Y, et al (2002) Influence of Different Cultivars on Populations of Ammonia-Oxidizing Bacteria in the Root Environment of Rice. Applied and Environmental Microbiology 68:3067–3075

Buddrus-Schiemann K, Schmid M, Schreiner K, et al (2010) Root Colonization by Pseudomonas sp. DSMZ 13134 and Impact on the Indigenous Rhizosphere Bacterial Community of Barley. Microbial Ecology 60:381–393

Carini P, Marsden PJ, Leff JW, et al (2017) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology 2:16242

Ciais P et al. in Climate Change 2013: The Physical Science Basis (eds Stocker, T. F. et al.) 465–570 (IPCC, Cambridge Univ. Press, 2013).

Carey JC, Tang J, Templer PH, et al (2016) Temperature response of soil respiration largely unaltered with experimental warming. Proceedings of the National Academy of Sciences 113:13797–13802

Cox PM, Pearson D, Booth BB, et al (2013) Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. Nature 494:341-344

D'Amico S, Collins T, Marx J-C, et al (2006) Psychrophilic microorganisms: challenges for life. EMBO reports 7:385–389

Dahl JL, Ulrich CH, Kroft TL (2011) Role of Phase Variation in the Resistance of Myxococcus xanthus Fruiting Bodies to Caenorhabditis elegans Predation. Journal of Bacteriology 193:5081–5089

Davidson EA, Janssens IA (2006) Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature 440:165–173

Delmont TO, Robe P, Cecillon S, et al (2011) Accessing the Soil Metagenome for Studies of Microbial Diversity. Applied and Environmental Microbiology 77:1315–1324

Dijkstra P, Thomas SC, Heinrich PL, et al (2011) Effect of temperature on metabolic activity of intact microbial communities: Evidence for altered metabolic pathway activity but not for increased maintenance respiration and reduced carbon use efficiency. Soil Biology and Biochemistry 43:2023–2031

Friedlingstein P, Meinshausen M, Arora VK, et al (2014) Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. Journal of Climate 27:511–526

Gregory JM, Jones CD, Cadule P, Friedlingstein P (2009) Quantifying Carbon Cycle Feedbacks. Journal of Climate 22:5232–5250

GISTEMP Team, (2018) GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies. Dataset accessed 2018-07-11

Hagerty SB, van Groenigen KJ, Allison SD, et al (2014) Accelerated microbial turnover but constant growth efficiency with warming in soil. Nature Climate Change 4:903–906

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Reviews of Geophysics 48:

Hespell RB, Rosson RA, Thomashow MF, Ritenberg SC (1973) Respiration of Bdellovibrio bacteriovorus strain 09J and its energy substrates for intraperiplasmic growth. Journal of Bacteriology 113:1280-1288

Höfle S, Rethemeyer J, Mueller CW, John S (2013) Organic matter composition and stabilization in a polygonal tundra soil of the Lena Delta. Biogeosciences 10:3145–3158

Intergovernmental Panel on Climate Change (2014) Evaluation of Climate Models. In: Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge, pp 741–866

Jenkinson DS, Adams DE, Wild A (1991) Model estimates of CO_2 emissions from soil in response to global warming. Nature 351:304–306

Jones MB, Highlander SK, Anderson EL, et al (2015) Library preparation methodology can influence genomic and functional predictions in human microbiome research. Proceedings of the National Academy of Sciences 112:14024–14029

Joos F, Spahni R (2008) Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years. Proceedings of the National Academy of Sciences 105:1425–1430

Jurkevitch E, Minz D, Ramati B, Barel G (2000) Prey Range Characterization, Ribotyping, and Diversity of Soil and Rhizosphere Bdellovibrio spp. Isolated on Phytopathogenic Bacteria. Applied and Environmental Microbiology 66(6):2365-2371

Kaiser C, Franklin O, Dieckmann U, Richter A (2014) Microbial community dynamics alleviate stoichiometric constraints during litter decay. Ecology Letters 17:680–690

Kallenbach CM, Frey SD, Grandy AS (2016) Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nature Communications 7:13630

Karhu K, Auffret MD, Dungait JAJ, et al (2014) Temperature sensitivity of soil respiration rates enhanced by microbial community response. Nature 513:81–84

Knutti R, Sedláček J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change 3:369–373

Lawrence DM, Slater AG, Swenson SC (2012) Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4. Journal of Climate 25:2207–2225.

Le Quéré C, Peters GP, Andres RJ, et al (2014) Global carbon budget 2013. Earth System Science Data 6:235–263

Liang C, Balser TC (2011) Microbial production of recalcitrant organic matter in global soils: implications for productivity and climate policy. Nature Reviews Microbiology 9:75–75

Liebner S, Wagner D (2007) Abundance, distribution and potential activity of methane oxidizing bacteria in permafrost soils from the Lena Delta, Siberia. Environmental Microbiology 9:107–117.

Lloyd J, Taylor JA (1994) On the Temperature Dependence of Soil Respiration. Functional Ecology 8:315

Luo Y, Ahlström A, Allison SD, et al (2016) Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles 30:40–56

Lüthi D, Le Floch M, Bereiter B, et al (2008) High-resolution carbon dioxide concentration record 650,000–800,000 years before present. Nature 453:379–382

Manzoni S, Taylor P, Richter A, et al (2012) Environmental and stoichiometric controls on microbial carbon-use efficiency in soils: Research review. New Phytologist 196:79–91

Nesme J, Achouak W, Agathos SN, et al (2016) Back to the Future of Soil Metagenomics. Frontiers in Microbiology 7

O'Connor K, Zusman D (1989) Patterns of cellular interactions during fruiting-body formation in Myxococcus xanthus. Journal of Bacteriology 171: 6013–6024

Paul EA (2016) The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. Soil Biology and Biochemistry 98:109–126

Pelini SL, Maran AM, Chen AR, et al (2015) Higher Trophic Levels Overwhelm Climate Change Impacts on Terrestrial Ecosystem Functioning. PLOS ONE 10:e0136344

Schuur EAG, Bockheim J, Canadell JG, et al (2008) Vulnerability of Permafrost Carbon to Climate Change: Implications for the Global Carbon Cycle. BioScience 58:701–714.

Schuur EAG, Vogel JG, Crummer KG, et al (2009) The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. Nature 459:556–559

Sierra CA, Müller M, Trumbore SE (2012) Models of soil organic matter decomposition: the SoilR package, version 1.0. Geoscientific Model Development 5:1045–1060

Sinsabaugh RL, Manzoni S, Moorhead DL, Richter A (2013) Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling. Ecology Letters 16:930–939

Sockett RE (2009) Predatory Lifestyle of Bdellovibrio bacteriovorus. Annual Review of Microbiology 63:523–539

Stocker TF et al in Climate Change 2013: The Physical Science Basis (eds Stocker, T. F. et al.) 33–115 (IPCC, Cambridge Univ. Press, 2013).

Tarnocai C, Canadell JG, Schuur EAG, et al (2009) Soil organic carbon pools in the northern circumpolar permafrost region. Global Biogeochemical Cycles 23:2

Todd-Brown KEO, Randerson JT, Post WM, et al (2013) Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10:1717–1736

Todd-Brown KEO, Randerson JT, Hopkins F, et al (2014) Changes in soil organic carbon storage predicted by Earth system models during the 21st century. Biogeosciences 11:2341–2356

Tripati AK, Roberts CD, Eagle RA (2009) Coupling of CO_2 and Ice Sheet Stability Over Major Climate Transitions of the Last 20 Million Years. Science 326:1394–1397

van Vliet S, Ackermann M (2015) Bacterial Ventures into Multicellularity: Collectivism through Individuality. PLOS Biology 13:e1002162

Waksman SA (1936) Humus, Origin, Chemical Composition and Importance in Nature (Williams and Wilkins, 1936).

Wang J, Sun J, Xia J, et al (2018) Soil and vegetation carbon turnover times from tropical to boreal forests. Functional Ecology 32:71–82.

Xia J, Luo Y, Wang Y-P, Hararuk O (2013) Traceable components of terrestrial carbon storage capacity in biogeochemical models. Global Change Biology 19:2104–2116

Xu X, Schimel JP, Janssens IA, et al (2017) Global pattern and controls of soil microbial metabolic quotient. Ecological Monographs 87:429–441

Zhou J, Xue K, Xie J, et al (2011) Microbial mediation of carbon-cycle feedbacks to climate warming. Nature Climate Change 2:106–110

Zhou X, Li S, Li W, et al (2014) Myxobacterial community is a predominant and highly diverse bacterial group in soil niches: Predominant soil myxobacteria. Environmental Microbiology Reports 6:45–56

Zickfeld K, Herrington T (2015) The time lag between a carbon dioxide emission and maximum warming increases with the size of the emission. Environmental Research Letters 10:031001

CHAPTER 2: PHOSPHATE ADDITION INCREASES TROPICAL FOREST SOIL RESPIRATION PRIMARILY BY DECONSTRAINING MICROBIAL POPULATION GROWTH

Eric R Johnston, Minjae Kim, Janet K Hatt, Jana R Phillips, Quiming Yao, Yang Song, Terry C Hazen, Melanie A Mayes, & Konstantinos T Konstantinidis

2.1. Abstract

Tropical ecosystems are an important sink for anthropogenic CO_2 emissions; however, sustained uptake is restricted by phosphorus (P) availability. Although soil microbiota facilitate organic P turnover and inorganic P mobilization, their role in tropical C-P-coupled biogeochemistry remains poorly understood. To advance this topic, soils collected from four sites with varying P content in El Yunque National Forest, Puerto Rico were incubated with exogenous PO_4^{3-} under controlled laboratory conditions. P amendment increased CO_2 respiration by 13.6-23.3% relative to control incubations for soils sampled from all but the most P-rich site. Metatranscriptomics revealed an increase in the relative transcription of genes involved in cell growth and uptake of other nutrients in response to P amendment. A methodology to normalize population-level gene expression by relative metagenome abundance revealed that P constraints on microbial growth were community-wide. Low-P-adapted communities were found to possess a greater relative abundance of α-glucosyl polysaccharide biosynthesis genes. Phosphorylase genes governing the

degradation of α-glucosyl polysaccharides were also more abundant in P-limited soils and increased in relative transcription with P amendment, indicating a shift from energy storage towards growth. Conversely, microbial communities inhabiting P-rich soils were found to have metabolisms tuned for the phosphorolysis of labile, plant-derived substrates (i.e., β-glucosyl polysaccharides). Collectively, our results provided quantitative estimates of increased soil respiration upon alleviation of P constraints and elucidated many underlying ecological and molecular mechanisms involved in this response.

2.2. Introduction

Despite covering <10% of the global land surface, tropical ecosystems account for one third of terrestrial net primary productivity and harbor one fourth of terrestrial biosphere carbon (C) (Bonan 2008). Tropical ecosystems have also played a major role in offsetting much of the anthropogenically-emitted carbon dioxide (CO_2) through enhanced plant uptake (Bala et al., 2007). However, current and sustained CO_2 uptake is limited by the availability of other essential nutrients, especially phosphorus (P) (Oren et al., 2001; Beedlow et al., 2004; Cleveland et al., 2011). In weathered tropical soils where inorganic P content is minimal or exists primarily in biologically inaccessible forms, P is often regarded as the primary limiting nutrient for biological activities (Vitousek 1984; Tanner et al., 1998). While several previous studies have focused on how P availability constrains aboveground communities and C cycling, microbial activities involved in P cycling and soil organic carbon (SOC) turnover (CO_2 release) remain much less understood (Turner and Wright 2014).

Soil P content is traditionally viewed as being dependent on the weathering of parent material (rock, minerals) over timescales involving

thousands of years. Phosphate molecules released by weathering are usually then readily immobilized into insoluble (and thus, biologically inaccessible) forms (Brady 1990). Soil organic P (SOP) molecules can comprise a large fraction of the total soil P, but their relative significance varies between soil types (Harrison 1987; Condron et al., 2005; Turner 2008). The release of inorganic phosphate from organic forms relies heavily on microbial enzymatic activities. Hence, belowground microbial activities are vital to the scavenging of inorganic P and for the turnover of P-containing plant and microbial detritus, thereby increasing P availability for both plants and microorganisms.

P content has been found to regulate microbial activities in tropical forests (Cleveland et al., 2002; Cleveland and Townsend 2006), and prior studies have shown that alleviating P constraints can result in enhanced soil microbial growth and activity. For instance, a decade-long P-fertilization experiment in the tropical forests of Panama significantly increased microbial biomass carbon (MBC), whereas fertilization with other nutrients (N, K, Mg, and Ca) had no significant effect on MBC (Turner and Wright 2014). While a 25% increase in fine litter fall from added P could have contributed to the observed increase in MBC (Wright et al., 2011), another study conducted at an adjacent site found no change in MBC when soil was supplemented with doubled plant litter inputs (Sayer et al., 2012). Therefore, the increased MBC in these studies is presumably attributable to direct microbial P usage. However, a mechanistic understanding of how the added P promoted microbial growth and which gene(s) can serve as biomarkers for this response remains elusive. Such biomarkers could be important for improved modeling and managing of P-limited tropical soil ecosystems.

Given that other nutrients are present in excess or are otherwise biologically obtainable, P-limitation could constrain microbial biomass directly by

limiting the *de-novo* biosynthesis of P-containing 'cell infrastructure components' (*e.g.,* DNA, RNA, glycerophospholipids, ATP, NADPH, etc.). Cultivated microorganisms adapted to P limitation are known to suppress their cell replication mechanisms while exhibiting a state of dormancy induced by P starvation (Drebot et al., 1990). During states where nutrients, *e.g.,* P, are limiting and organic matter substrates are available in excess, microbiota have been shown to biosynthesize and accumulate α-glucosyl polymers (such as glycogen, trehalose, etc.) to serve as readily accessible energy sources for when nutrient conditions improve (Zevenhuizen 1966; Lilie and Pringle 1980). However, it remains unclear if the previous findings based mostly on lab-scale experiments with microbial isolates apply to *in-situ* soil conditions involving highly diverse substrates and complex traits of microbial communities, or what other traits low-P-adapted soil microorganisms possess.

The influence of P on microbial SOC mineralization is also poorly understood. Heightened microbial biomass sustained by abundant P resources could increase heterotrophic SOC demand, resulting in greater overall respiration. However, increased P content could also decrease activities involved in P scavenging and turnover. For instance, previous enzyme assay analyses conducted on P-limited tropical soils have shown that phosphoesterase activity is approximately one order of magnitude greater than those of SOC-degrading enzymes and that long-term P fertilization substantially decreases phosphoesterase activity (Turner and Wright 2014; Yao et al., 2018). These results suggest that P-limited microorganisms devote a large amount of resources towards P scavenging. Thus, P availability likely mediates, to an extent, the content and composition of organic matter in P-limited tropical forest ecosystems through differences in SOP turnover or catabolic and anabolic SOC

usage. Yet, many of the corresponding physiological and molecular mechanisms, *e.g.,* differences in gene content and transcription related to SOC degradation or biomass formation, remain poorly understood. To improve modeling of SOC stocks and CO_2 emissions in Earth system models and to more accurately quantify the future effects and feedbacks of global change factors on tropical ecosystems, the mechanistic representation of coupled C-P biochemistry is essential (Cavaleri et al., 2015; Holm et al., 2017).

In this study, we evaluated how low-P and P-rich tropical soils differ in their microbial community composition and functioning, and how these communities respond to an alleviation of prior nutrient constraints (through the addition of a readily available P source). For this, soils were collected from four locations in the El Yunque National Forest in Puerto Rico with varying *in-situ* P content. Short-term laboratory incubation experiments were employed to relate heightened soil respiration from the addition of exogenous phosphate (P amendment) to changes in microbial community structure and transcriptional activity (assessed by metagenomics and metatranscriptomics), and to enzyme activities involving SOC and SOP mineralization. Our goal was to address the following hypotheses: 1) microbial communities in P-limited soils will possess traits indicative of such - *e.g.,* more abundant, diverse, and active mechanisms of P acquisition or pathways for the retention of organic matter substrates, 2) heightened soil respiration resulting from P amendment will be relatable to prior *in-situ* P limitation and to increased expression of genes involved in central anabolic and catabolic activities, and 3) multi-omics approaches integrating metagenomic and metatranscriptomic datasets (*e.g.,* normalizing transcriptional activity to *in-situ* DNA abundance) can permit more comprehensive and quantitatively resolved microbial responses to experimental P amendment.

2.3. Materials and methods

Methods for soil physiochemical determinations and enzyme assays can be found in the supplemental material.

2.3.1. Study site description, sample collection, and sample transport

The Luquillo Experimental Forest within the El Yunque National Forest in northeastern Puerto Rico is classified as a subtropical wet forest. Oxisol soils weathered from volcaniclastic sediments were collected from 350 masl at the El Verde Field Station (18°19'16"N, 65°49'12"W) near the town of Río Grande (Silver et al., 1999; Mage and Porder 2013). Ultisol soils weathered from volcaniclastic sandstones were collected from 400 masl from the Bisley Experimental Watershed near the Sabana Field Research Station (18°19'32"N, 65°43'44"W) and the town of Luquillo (McGroddy and Silver 2000; Liptzin and Silver 2009; Wood et al., 2016). Both locations are associated with the Luquillo Long-term Ecological Research network and the Luquillo Critical Zone Observatory. The sites receive around 3500mm of rainfall per year and have a mean annual temperature of ~23°C. Common forest tree species include Prestoea Montana R. Graham Nichols (palm) and Dacryodes excelsa Vahl (tabonuco).

The topography consists of numerous small-scale ridges and valleys (Scatena 1989; Mage and Porder 2013). At El Verde (E) and at Bisley (B) watersheds, samples were collected from valley (V) and ridgetop (R) locations on January 12-13, 2016. Triplicate soil samplings of ~1kg were taken at a depth of 10cm using a small trowel. The soils were shipped overnight on blue ice to

ORNL, where they were stored at 4°C until the start of the incubation experiments (<2 weeks after receipt).

2.3.2. Incubation procedure

For greater experimental replication and to incubate soils promptly after sampling, incubation experiments were conducted on two automated MicroOxymax Respirometer Systems (Columbus Instruments, Columbus, OH) capable of monitoring CO_2, including a 20-flask capacity system at Oak Ridge National Laboratory (ORNL) and a 30-flask capacity system located at the University of Tennessee at Knoxville (UT). For each incubation, ~100g of wet soil was added to a sterilized 250mL incubation vessel. Three mL of sodium phosphate solution (equivalent to 30μg P g soil^{-1}) was evenly distributed across incubation soils; 3mL of molecular grade water was evenly distributed across a separate incubation to serve as control. This readily available form of P was used in order to study the straightforward and immediate response of soils to an alleviation of prior nutrient constraints. Incubations were held at 27ºC, reflecting the average daily temperature in the nearby city of San Juan, Puerto Rico from May-October (The Southeast Regional Climate Center; http://www.sercc.com). The first incubation experiment (conducted at ORNL) lasted for ~6 days and included 2 of the 3 triplicate field samplings from each of the four sites to compare respiration under control and P-amendment conditions (16 incubations) (Figure A-2). Two autoclaved soils and two empty jars were incorporated to ensure that measured gases resulted from biological activity. Gas measurements were logged every ~2.16 hours. A second incubation experiment was performed similarly (conducted at UT), but included all three replicate field samplings (24 incubations) and lasted ~14 days and gas measurements were logged every

~2.96 hours. At the conclusion of each experiment, soil subsamples from each jar were immediately subjected to liquid nitrogen prior to storage at -80ºC in order to preserve soil community RNA.

2.3.3. Soil RNA and DNA isolation, library preparation, and sequencing

Soil DNA and total RNA was simultaneously extracted (from samples demarcated in Figure A-2) using the RNA PowerSoil Total RNA Isolation Kit and the RNA PowerSoil DNA Elution Accessory Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) according to manufacturer's protocol, using ~0.5 g of soil for each extraction. Total RNA samples were treated with DNase using the TURBO DNA-free kit (Ambion, Austin, TX). Elimination of DNA from RNA samples was confirmed by PCR amplification of the 16S rRNA gene using universal primers 8F and 1492R, followed by gel-electrophoresis. RNA integrity was assessed with an Agilent 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA), using the Agilent RNA 6000 Pico Kit. Generally, total RNA samples with a RIN (RNA integrity number) of <7 (as determined by Agilent Bioanalyzer software) were discarded. Following DNA digestion and evaluation of RNA integrity, two to four replicate extractions were pooled for each sample in order to obtain enough high quality RNA and to overcome high sample-to-sample heterogeneity, which is characteristic of soil microbial communities. Ribosomal RNA (rRNA) was depleted from total RNA using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina, San Diego, CA) and cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, San Diego, CA) according to manufacturer's instructions, but with two modifications: RNA was fragmented for 1 minute instead of 5 in step 3.A, and AMPure XP purification

(step 3.F.1) was performed twice in order to reduce the amount of primer-dimers generated during amplification of cDNA.

The quality of each library was inspected on an Agilent 2100 Bioanalyzer, using the Agilent High Sensitivity DNA kit. RNA and DNA quantification were achieved using Qubit RNA HS Assay Kit and Qubit dsDNA HS Assay Kit, respectively (Thermo Fisher Scientific). Dual-indexed DNA sequencing libraries were prepared using the Illumina Nextera XT DNA library prep kit according to manufacturer's instruction except the protocol was terminated after isolation of cleaned double-stranded libraries. Library concentrations were determined by fluorescent quantification using a Qubit HS DNA kit and Qubit 2.0 fluorometer (ThermoFisher Scientific) according to the manufacturer's instructions and samples were run on a High Sensitivity DNA chip using the Bioanalyzer 2100 instrument (Agilent) to determine library insert sizes. DNA and cDNA libraries were sequenced at the Georgia Institute of Technology High Throughput DNA Sequencing Core on an Illumina HiSeq 2500 instrument in the rapid run mode for 300 cycles (150 bp paired-end). Additional sequencing of the DNA libraries was performed on an Illumina NextSeq 500 instrument, also in the high output mode for 300 cycles (150 bp paired-end). Adapter trimming and demultiplexing of sequences was carried out by either Illumina instrument. Metagenomic and metatranscriptomic datasets were deposited in the European Nucleotide Archive under project PRJEB23349.

2.3.4. Bioinformatics data analysis

Both metagenomic and metatranscriptomic paired-end reads were merged using PEAR (Zhang et al., 2014) (options: -p 0.001). All merged and non-merged reads were then quality-trimmed with the SolexaQA package (Cox et al., 2010)

(options: -h 17; >98% accuracy per nucleotide position). Trimmed sequences used for downstream functional annotation were truncated to 150 bp to avoid read-length biases. SortMeRNA was used with default settings to identify and remove residual rRNA sequences from sample metatranscriptomes (i.e., those remaining after rRNA subtraction) (Kopylova et al., 2012). Protein-prediction for short-reads was performed with FragGeneScan (Rho et al., 2010) (Illumina 1% error model). Resulting amino acid sequences were searched against Swiss-Prot database (UniProt Consortium 2015; downloaded on 11-27-2016), using blastp (BLAST+ version 2.2.28) (options: -word_size 4, cutoff: bit score ≥55) (Camacho et al., 2009). Corresponding gene and enzyme accession IDs for each Swiss-Prot entry was obtained from http://www.uniprot.org/downloads (downloaded on February 28, 2016) to consolidate the annotations. Taxonomic composition was determined using 16S rRNA gene fragments recovered from metagenomes using SortMeRNA (Kopylova et al., 2012), which were annotated and summarized at the phylum-level as described previously (Johnston et al., 2016). Metagenomic datasets belonging to the same site (*e.g.,* Bisley Valley) were combined for co-assembly (combining data from multiple sample metagenomes to increase coverage) with idba_ud (Peng et al., 2012) (options: --mink 55 --maxk 123 --step 4). Resulting contigs ≥1 kbp were retained for further analysis. Protein-encoding genes of the assembled contigs were predicted with Prodigal (Hyatt et al., 2010) and were searched against the Swiss-Prot and TrEMBL database as described above (UniProt Consortium 2015).

Metatranscriptomes were checked for DNA contamination by mapping short-read transcripts to assembled contigs and determining strand-specificity (consistency in sense/antisense orientation) for the 1,000 most transcriptionally-active genes. Samples with <95% average strand-specificity were not used for

analysis; one sample metatranscriptome, El Verde ridge 1 control incubation, failed this criterion.

To evaluate transcriptional activity of cell division and cell wall operons (*dcw*) represented by distinct bacterial populations, the following methodology for integrating metagenome and metatranscriptome datasets representing the same subsamples was used. First, assembled contigs (from DNA) encoding ≥5 sequential genes of the *dcw* operon, including genes *ftsAIQWYZ*, *murBCDEFG*, *mraY*, and *ddl*, were identified. In an attempt to normalize population-specific transcriptional activity to *in-situ* (DNA) abundance, *i.e.,* a population could show differential transcript abundance between samples due to more/fewer cells sampled and not necessarily increased/deceased transcription, the transcriptional coverage of each individual operon was divided by its DNA sequence coverage. Metagenome and metatranscriptome coverage (*i.e.* the mean representation of a contig by matching short-reads) was determined using MegaBLAST (Camacho et al., 2009), retaining only reads that matched at ≥98% nucleotide identity, ≥100bp match, and ≥90% of the short-read length. Relatedness between individual *dcw* operons was determined by aligning full-length assembled and reference *murG* amino acid sequences (obtained from Uniprot.org) using MAFFT (Katoh and Standley 2013). Phylogenetic distances from the resulting alignment were determined using PhyML 3.0 (Guindon et al., 2010). Visualization of the resulting *murG* sequence-based phylogenetic tree(s), coupled to the DNA-normalized transcriptional activity, were generated using the web-based software iTOL (Letunic and Bork 2016).

2.3.5. Statistical analyses

Paired t-test was performed to evaluate the significance of heightened CO_2 respiration between control and P-amended incubations for all four sampling locations, individually. ANOVA with Tukey's HSD was used to distinguish sampling locations that differed significantly by soil physiochemical indices, enzyme activity measurements, and the relative abundances of microbial phyla. Further, paired t-test was used to identify significant enzyme activity differences between control and P-amended soils from Bisley Ridge and El Verde Ridge. Tables with raw counts of genes (based on Swiss-Prot annotations; UniProt Consortium 2015) were processed with the edgeR software package (Robinson et al., 2010) to identify significant, differentially abundant genes between sample groups (ridge vs. valley metagenomes) or treatments (control vs. P-amended incubation metatranscriptomes). To better assess transcriptional differences resulting from P amendment, a paired design was adopted so that baseline differences between sampling sites (i.e., Bisley Ridge vs. El Verde Ridge) were subtracted out. P-values from metagenome edgeR analysis were transformed to account for false discovery rate using Benjamini–Hochberg correction (Benjamini and Hochberg 1995). P-value correction was not performed for metatranscriptome results due to the small number of datasets used for comparison (primarily due to the difficulty in obtaining high quality RNA from soils). To account for the latter, our analysis was focused on inspecting for consistent transcriptome responses between multiple genes involved in the same pathway, operon, or functional category (which showed a high degree of consistency, as demonstrated in the results section). Functional annotations with an average relative abundance <1E-6 for metagenomes and <1E-5 for

metatranscriptomes were not considered for comparisons between sample groups.

2.4. Results

2.4.1. Relationships between environmental indices and CO_2 respiration during incubation

A summary of soil chemical and physical measurements is provided in Table 2-1. Available phosphorus (extractable P from Mehlich-1 method) varied significantly between study locations (ANOVA, adj. p-value<0.05); it was lowest in Bisley Ridge (BR) and El Verde Valley (EV) soils at 0.9±0.1 and 0.8±0.3mg P kg dry soil^{-1}, respectively, was 1.9±0.1mg P kg^{-1} in El Verde Ridge (ER) soils, and was highest in Bisley Valley (BV) soils at 3.0±0.3mg P kg^{-1}.

P addition increased CO_2 respiration for EV, ER, and BR soils significantly (+13.6%, +13.8%, and +23.3% mean increase from replicate incubations, respectively; p-value<0.05), and resulted in a smaller, non-significant increase for BV soils (+4.6%) (Figure 2-1). A positive correlation between microbial biomass phosphorus (MBP) and available P was observed in field samples from ridge soils (R^2=0.82) (Figure A-3a). Control soils with higher MBP respired more CO_2 after 14 days (R^2=0.80) (Figure A-3b). Control soils with a higher MBC:MBP ratio respired less CO_2 (R^2=0.85) (Figure A-3c). The increase in CO_2 respiration with P amendment (percent increase vs. control incubation) was positively correlated with MBC:MBP ratio (R^2=0.86) (Figure A-3d). Positive correlations were also observed between MBP and both MBC (R^2=0.77) and MBN (R^2=0.60) (Figure A-3ef).

Table 2-1: Summary of soil physiochemical indices used to characterize the four sampling sites. Values are given as the mean ± the standard error, derived from measurements on triplicate samplings. Superscript letters are used to distinguish sample groups that were significantly different (adj. p-value<0.05), *i.e.,* values with superscript letters differing from letters assigned to other values designates a statistically-significant difference between soils for that measurement, as determined with ANOVA in conjunction with Tukey's HSD test.

Study Site	pH	Soil Moisture %	NH4-N (mg/kg)	NO3-N (mg/kg)	Total P (mg/kg)	Total C (g/kg)	Total N (g/kg)
El Verde Ridge	4.65 ± 0.08 ^{a}	0.36 ± 0.01 ^{a}	43.8 ± 10.8 ^{ab}	6.99 ± 0.55 ^{a}	334.0 ± 9.08 ^{b}	6.64 ± 0.32 ^{b}	0.50 ± 0.02 ^{c}
Bisley Ridge	4.82 ± 0.12 ^{a}	0.41 ± 0.02 ^{a}	65.5 ± 18.2 ^{b}	1.23 ± 0.97 ^{a}	206.3 ± 14.5 ^{a}	5.33 ± 0.42 ^{b}	0.38 ± 0.03 ^{b}
El Verde Valley	5.41 ± 0.06 ^{b}	0.45 ± 0.04 ^{a}	10.3 ± 1.1 ^{a}	9.21 ± 3.50 ^{a}	364.5 ± 27.8 ^{b}	3.07 ± 0.35 ^{a}	0.25 ± 0.03 ^{a}
Bisley Valley	5.40 ± 0.17 ^{b}	0.45 ± 0.02 ^{a}	37.4 ± 9.0 ^{ab}	11.4 ± 8.74 ^{a}	447.8 ± 15.0 ^{c}	5.24 ± 0.42 ^{b}	0.39 ± 0.02 ^{b}

	Measurements from Mehlich-1 (mg/kg dry soil)					
Study Site	Ca	K	Mg	Mn	Zn	P
El Verde Ridge	136.5 ± 17.0 ^{a}	45.6 ± 3.7 ^{b}	121.0 ± 11.4 ^{a}	18.4 ± 4.7 ^{a}	4.5 ± 1.3 ^{a}	1.9 ± 0.1 ^{b}
Bisley Ridge	266.6 ± 78.0 ^{a}	16.7 ± 0.5 ^{a}	205.3 ± 26.5 ^{b}	42.4 ± 25.8 ^{a}	2.1 ± 0.8 ^{a}	0.9 ± 0.1 ^{a}
El Verde Valley	652.3 ± 31.6 ^{b}	40.7 ± 5.1 ^{b}	356.0 ± 21.9 ^{c}	41.2 ± 8.9 ^{a}	1.8 ± 0.2 ^{a}	0.8 ± 0.3 ^{a}
Bisley Valley	842.2 ± 93.7 ^{b}	15.6 ± 4.2 ^{a}	301.9 ± 6.3 ^{c}	35.5 ± 6.9 ^{a}	2.4 ± 0.1 ^{a}	3.0 ± 0.3 ^{c}

	Microbial biomass indices			Soil texture analysis		
Study Site	MBC (mg/g dry soil)	MBP (ug/g dry soil)	MBN (mg/g dry soil)	Sand (%)	Clay (%)	Silt (%)
El Verde Ridge	1.78 ± 0.19 ^{c}	5.59 ± 0.54 ^{b}	0.28 ± 0.04 ^{b}	16.5 ± 5.9 ^{a}	49.1 ± 3.1 ^{b}	34.5 ± 3.6 ^{a}
Bisley Ridge	1.05 ± 0.16 ^{b}	2.26 ± 0.50 ^{a}	0.18 ± 0.03 ^{ab}	3.5 ± 1.8 ^{a}	63.4 ± 1.6 ^{c}	33.5 ± 2.8 ^{a}
El Verde Valley	0.44 ± 0.02 ^{a}	0.23 ± 0.07 ^{a}	0.08 ± 0.00 ^{a}	7.0 ± 4.4 ^{a}	42.1 ± 3.7 ^{ab}	50.9 ± 1.1 ^{b}
Bisley Valley	0.61 ± 0.09 ^{ab}	1.80 ± 0.60 ^{a}	0.10 ± 0.02 ^{a}	20.1 ± 2.8 ^{a}	33.9 ± 1.6 ^{a}	46.1 ± 1.6 ^{b}

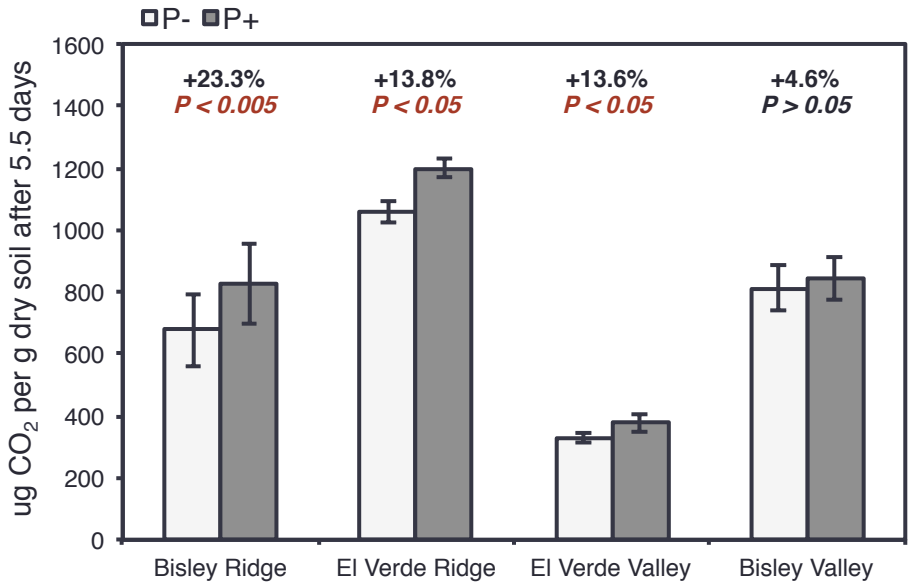


Figure 2-1: Cumulative CO_2 respired after ~5.5 days of incubation for soils without (P-, light gray) and with (P+, dark gray) exogenously added phosphorus. Bars denote the mean CO_2 respired (+/- the standard error of the mean) for control incubations vs. incubations amended with exogeneous phosphate as 30 mg PO_4^{3-}-P per kg for soils sampled from each of the four locations.

2.4.2. Organic carbon degrading and phosphoesterase enzyme activities

Enzyme assays and metagenomic and metatranscriptomic sequencing efforts focused on incubations representing low-P ridge (i.e., ER and BR) and P-rich BV soils due to the difficulty in obtaining high quality RNA from EV soils. See Figure A-2 for a diagram of incubations performed and those targeted for enzyme assays and RNA and DNA extraction and sequencing.

Enzyme activities were evaluated as a ratio between SOC-degrading enzyme activities (namely, those of α-glucosidase, β-glucosidase, β-D-cellubiosidase, N-acetyl β-glucosaminidase, and β-xylosidase) and phosphodiesterase (DIPHOS) activity (*e.g.*, β-glucosidase:DIPHOS), an approach used previously to compare enzyme activities between soils (Sinsabaugh et al., 2009; Ramirez et al., 2012; Turner and Wright 2014). DIPHOS activity was ~2.5- to ~54-fold greater than that of SOC-degrading enzymes in low-P ridge soils (Figure A-4). Activity ratios were significantly greater for all five SOC-degrading enzyme classes for P-rich BV soils compared to ridge soils (~4.5 times greater, on average; adj. p-value<0.05 ANOVA). P-amended ridge incubation soils demonstrated a significant increase in relative β-D-cellubiosidase activity (+21.9%; p-value<0.05, paired t-test) and near-significant increases in α-glucosidase (+46.0%; p-value<0.1) and β-glucosidase (+5.3%; p-value<0.1) activities compared to field and control incubation soils

2.4.3. Broad compositional characteristics differentiating soil communities

Ten metagenome libraries representing ER, BR, and BV incubation soils and eight metranscriptomes, *i.e.,* four from low-P ridge (i.e., ER and BR) control incubations and four from the corresponding P-amended incubations, were sequenced to relate community composition and transcriptional differences with enzyme activities, respiration activity/response to P amendment, and soil physiochemical measurements. On average, 5.2Gbp of raw sequencing effort was obtained for each metatranscriptome and 16.3Gbp for each metagenome. Following paired-end read merging, quality trimming, and filtering of rRNA sequences (for metatranscriptomes), 2.0Gbp of metatranscriptomic and 12.2Gbp of metagenomic data was used for downstream analyses.

Using Nonpareil 3 (options: -T kmer -X 100000) (Rodriguez-R et al., 2018), P-rich BV soil communities were found to be more diverse than low-P ridge soils (i.e., ER and BR) (Figure A-5). The estimated sequencing depth needed to achieve 75% average coverage for ER, BR, and BV metagenomes was 14.3, 15.0, and 23.4Gbp, respectively. The level of coverage obtained (~40-60%) is appropriate for comparisons (e.g. low false negative rate), with un-sampled diversity presumably representing comparatively rarer taxa (Rodriguez-R and Konstantinidis 2014; Zhang et al., 2017). The composition of bacterial phyla derived from metagenomic 16S rRNA gene fragments was comparable between ridge locations, and more distinct between ridge and BV soils (Figure A-6; Figure A-7b). Considering phyla that represented $\geq$1% of the community for at least one metagenome, only *Verrucomicrobia* abundance differed significantly between ER and BR (10.4% vs. 5.6% for ER and BR soils, respectively; adj. p-value<0.05). In contrast, the abundance of several phyla differed significantly between low-P ridge and P-rich BV soils, including *Chloroflexi* (2.2% vs. 7.2% for ridge and BV soils, respectively), *Firmicutes* (1.0% vs. 0.5%), *Gemmatimonadetes* (0.17% vs. 1.3%), *Latescibacteria* (0% vs. 1.4%), *Nitrospirae* (0.1% vs. 4.3%), *Planctomycetes* (8.2% vs. 6.8%), and *Proteobacteria* (40.4% vs. 35.9%). Mash, a tool that uses kmer composition for β-diversity calculations (Ondov et al., 2016), also revealed greater similarity between ER and BR, than between ER/BR and BV (Figure A-7a).

Analyses of community composition revealed high overlap, in general, between P-amended and control incubation metagenomes (Figure 2-2; Figure A-6; Figure A-7). Thus, comparisons between control and P-amended communities were limited to metatranscriptomic data analyses.

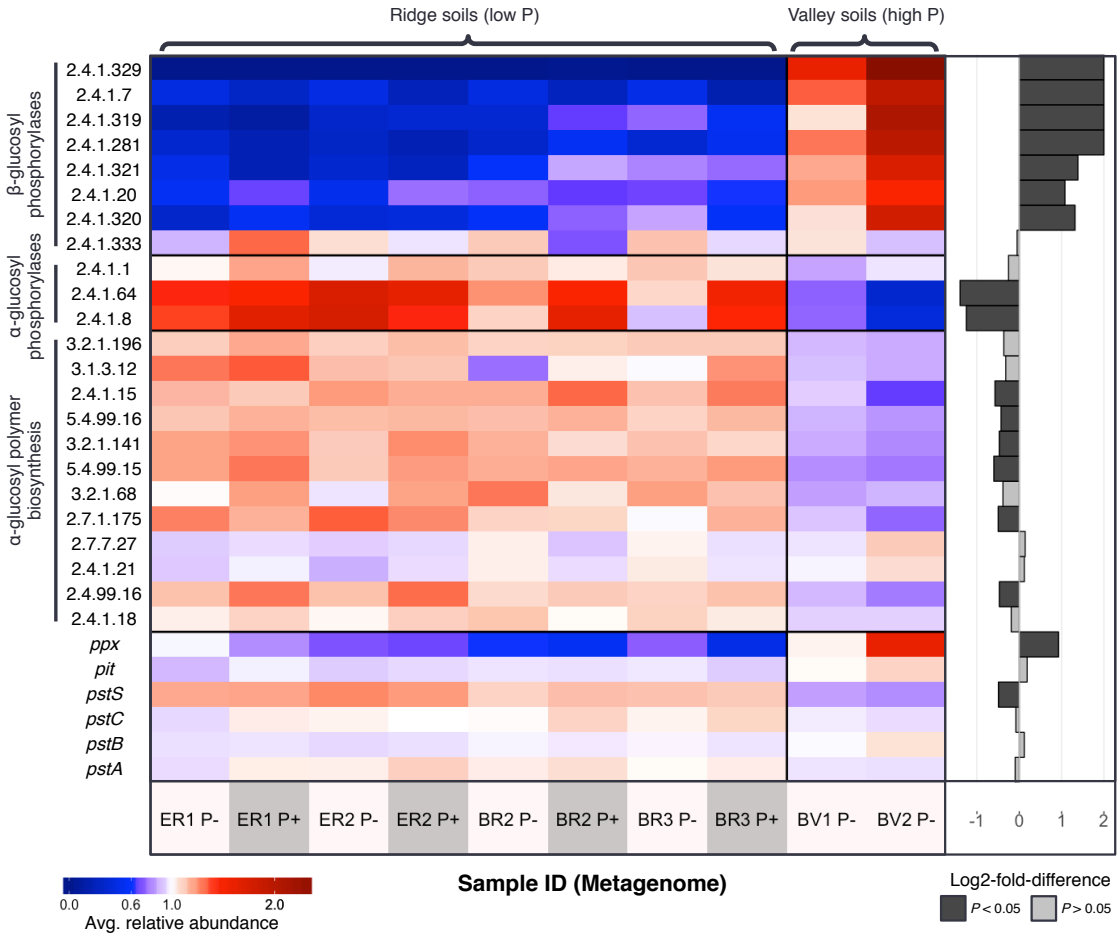


Figure 2-2: Heatmap showing the relative DNA abundances of genes involved in P uptake and storage, α-glucosyl polymer synthesis, and phosphorolysis of α- and β-glucosyl substrates. To emphasize differences between datasets, values were normalized by the average metagenome-derived gene abundance of all datasets (by rows). Sub-plots on the right of each heatmap represent the log2-fold-difference calculated for each gene, and are colored by whether differences between valley and ridge soil metagenomes were statitistical significant (see figure key). A log2-fold-difference of +2 was set as the maximum value for the right panel (some values from enzyme IDs under 'β-glucosyl specific phosphorylases were greater than +2).

2.4.4. Abundance and transcriptional activity of α-glucosyl polymer biosynthesis, anabolic and catabolic process, and cell division genes

The relative transcription of genes encoded by the universal *dcw* operon generally increased with P-amendment compared to control incubation samples (Figure 2-3). A multi-omics approach normalizing transcriptional activity by *in-situ* population abundance and metatranscriptome dataset size demonstrated that heightened transcription of genes for microbial cell division and growth with P amendment was community-wide (Figure 2-4). Specifically, >85% of *dcw* operons had greater DNA-normalized relative transcription under P-amendment conditions (vs. control); the relative total transcription of all recovered *dcw* segments was 56.4% and 87.0% greater with P-amendment (vs. control) for ER and BR soils, respectively. Genes involved in the biosynthesis of purines (*purABCDEFHKMT*, *guaAB*), pyrimidines (*carAB*, *pyr1BCDEFHG*, *ndk*, *nrdAB*), and glycerophospholipids (*tpiA*, *gspA*, *plsBCXY*, *pls*, *cdsA*) also generally increased in relative transcription for P-amended incubations (Figure 2-3).

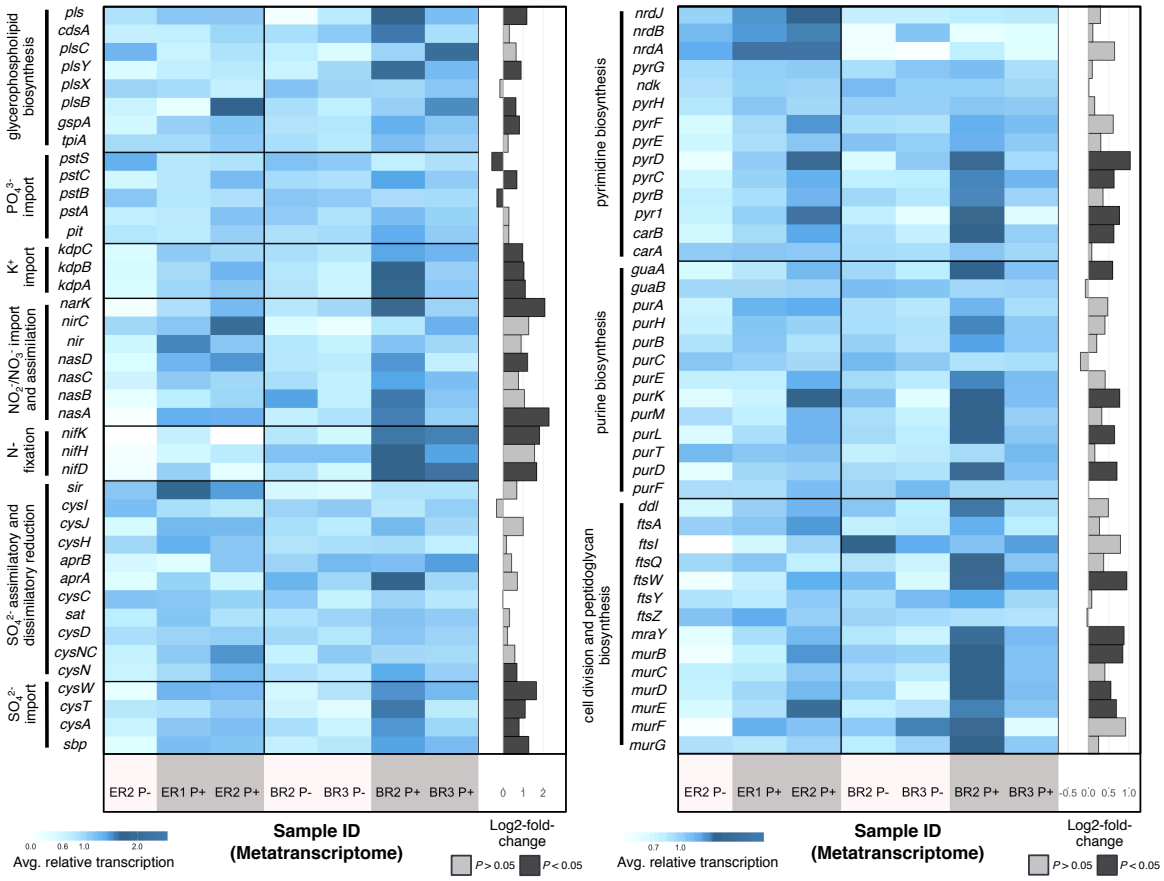


Figure 2-3: Heatmap showing the relative transcriptal activity of genes involved in nutrient uptake and assimilation, cell division, and cell component biosynthesis. Columns are represented by control (P-) and P-amended (P+) incubation metatranscriptomes from El Verde ridge (ER) and Bisley ridge (BR) soils. Values were normalized by the mean relative transcript abundance of all datasets (by rows). Sub-plots on the right of each heatmap represent the log2-fold-change calculated for each gene, and are colored by statitistical significant differences between P+ and P- soil incubation metatranscriptomes (see figure key).

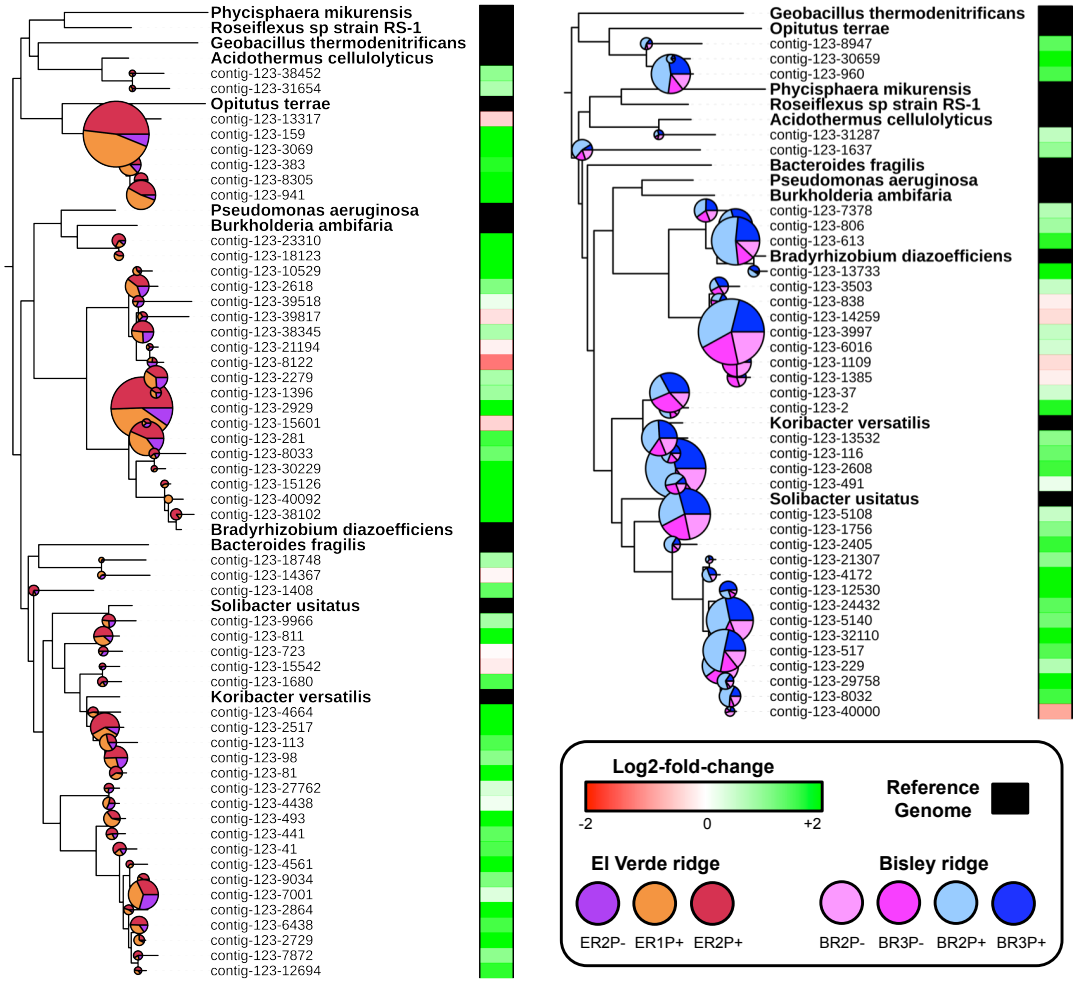


Figure 2-4: Population-specific transcriptional activity of the common cell division and cell wall (*dcw*) operon for El Verde ridge (left) and Bisley ridge (right) community co-assemblies. The phylogenetic tree was constructed using aligned *murG* sequences. Pie graphs represent the relative DNA-normalized transcriptional activity between control (P-) and P-amended incubation samples (P+) for each unique *dcw* operon recovered (each branch), and the right box (green/red) represents the log2-fold difference in activity between P- and P+ samples for each sequence/population. A log2-fold change of +2 was set as the maximum value (some positive values were slightly greater than +2).

The relative DNA abundances of genes for α-glucosyl polymer biosynthesis and phosphorylase genes acting on α-glucosyl polymers were generally greater in ridge metagenomes compared to BV (Figure 2-2). The opposite trend was found for phosphorylase genes specific to β-glucosyl polymers. The relative transcription of all phosphorylases acting on carbohydrates increased with P amendment (Figure A-8). A survey of metagenome-derived contigs revealed that many microbiota possessed genomic segments encoding genes for the debranching and phosphorolysis of large, α-glucosyl polymers, *i.e.*, glycogen debranching enzyme (EC 3.2.1.196) and glycogen/maltodextrin phosphorylase (EC 2.4.1.1). These were often adjacent to genes involved in core catabolic functions and/or the formation of several biosynthetic precursors relevant to cell growth, *e.g.*, genes involved in the formation of D-ribose 5-P (for nucleotide and ATP biosynthesis), D-erythrose 4-P (for amino acid biosynthesis), and glycerone phosphate (for phospholipid biosynthesis) (Figure 2-5). Genomic segments possessing $\geq$5 sequential genes related to glycogenolysis, (*glgX*, *malQ*, *pgm*, and phosphorylases specific to glycogen/maltodextrin [EC 2.4.4.1]), the initial stages of glycolysis (*pgi*, *pfkA*, and *fba*), and/or the pentose phosphate pathway (*pgl*, *gnd*, *rpe*, *rpiA*/*ywlf*, *tkt*, and *tal*) were evaluated for changes in transcriptional activity with P-amendment (vs. control). As an additional stipulation, only segments with $\geq$2 genes belonging to the glycogenolysis category were retained. P amendment increased the transcriptional coverage of these recovered segments by +250% and +103%, on average, for ER and BR soil incubations communities, respectively.

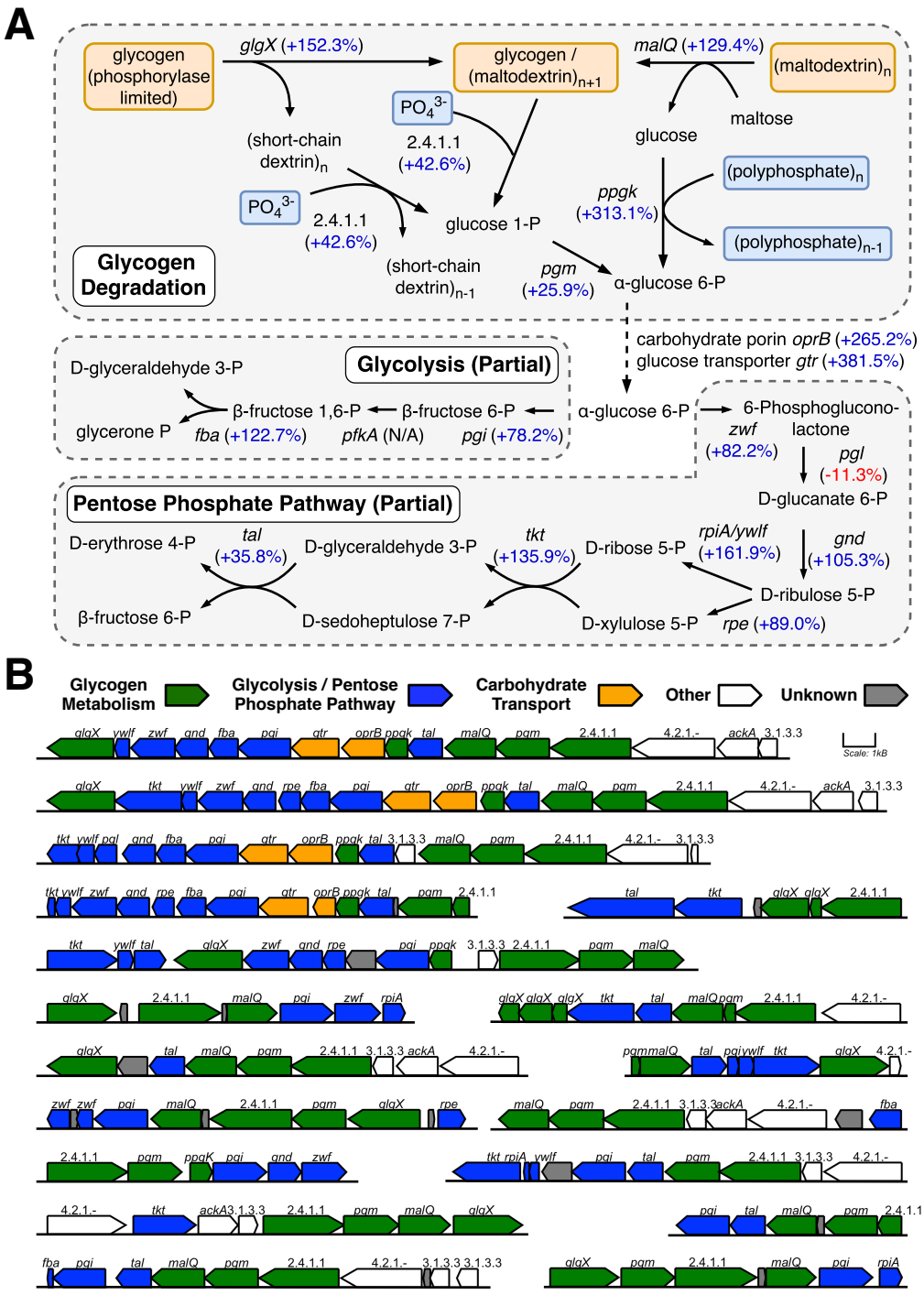


Figure 2-5: (A) Conceptual pathway diagram and (B) gene organization reflecting genomic segments involved in glycogen metabolism, glycolysis, and the pentose phosphate pathway. Percentages given in blue or red

represent the average increase or decrease in the relative transcription of the corresponding genes, respectively. Gene tracks represent unique contigs assembled from Bisley and El Verde ridge metagenomes that encoded genes of interest. Gene Graphics (Harrison et al., 2017) was used to produce 'gene track' illustrations. Continuous black lines under the gene boxes are used to signify unique, continuous genomic segments.

Genes involved in electron transport chain processes, such as bacterial cytochrome c-type biogenesis (*ccmABCEFH*), cytochrome bd-II ubiquinol oxidase (*cydAB*), cytochrome bo(3) ubiquinol oxidase (*cyoABC*), succinate dehydrogenase (*sdhAB*), and fumarate reductase (*frdAB*), generally increased in relative transcription for P-amended incubation samples (Figure A-9). While genes encoding for the hydrophobic membrane fragment NADH dehydrogenase (*nuoAHJKLMN*) showed a marginal but non-significant increase in relative transcription, the relative transcription of genes for other components of NADH dehydrogenase were similar between P-amended and control incubation samples.

2.4.5. Abundance and transcriptional activity of nutrient acquisition and storage genes

While the relative DNA abundances of genes encoding for phosphate transport ATPase complex (*pstABC*) were not significantly different between sites, that of *pstS* (encoding phosphate-binding protein) was 42.6% greater in low-P ridge incubation metagenomes relative to P-rich BV (Figure 2-2; adj. p-value<0.05). *Pit* (encoding low-affinity phosphate transporter) was more

abundant in P-rich BV than ridge soils, albeit non-significantly. The relative abundance of *ppx* (encoding exopolyphosphatase), a polyphosphatase with strong preference for long-chain polyphosphates, was 89.7% greater in P-rich BV incubation metagenomes (adj. p-value<0.05).

Consistent with the metagenome differences between high and low P soils, the relative transcription of *pit* was greater (+20.7%) in P-amended ridge incubation soils (vs. control), albeit non-significantly. The phosphate ABC transporter membrane subunit genes (*pstAC*) increased in relative transcription with P amendment (Figure 2-3), but *pstB* and *pstS* (encoding ATP-binding protein) decreased (p-value<0.05). Genes involved in the uptake, fixation, and/or assimilation other nutrients generally increased in relative transcription with P amendment. This included genes involved in N fixation (*nifDHK*), ammonium import (*nrgA*), assimilatory nitrate reduction (*nasABCD*, *nir*), sulfur binding and import (*cysATW*, *sbp*), and potassium import (*kdpABC*). The relative transcription of genes for assimilatory sulfate reduction (*cysCDNJ*, *sat*, *sir*) also generally increased with P amendment, but these differences were comparably minor and non-significant.

2.4.6. Transcriptional activity of cellular stress and maintenance genes

The relative transcription of *frr* (ribosomal recycling factor) and nearly every ribosomal protein gene was significantly lower in P-amended ridge incubation soils (Figure A-9; p-value<0.05). The relative transcription of genes related to cellular stress or maintenance, such as for protecting against reactive oxygen species (*prdx*, *grx*, *sodCFMN*), protein degradation (*lon, clpACPS*), molecular chaperones (*CH60*, *CH10*, *clpB*, *dnaKJ*), and cold or heat-shock

proteins (*cpsABCE*, *hspBCDH*, *idpA*), were lower in P-amended soil metatranscriptomes.

2.5. Discussion

Phosphorus is often regarded as the primary limiting nutrient for biological activities in tropical ecosystems (Vitousek 1984; Tanner et al., 1998), yet many uncertainties remain concerning how P availability governs tropical soil microbial community composition and activities (Turner and Wright 2014). In this study, P amendment of lab-incubated soils increased CO_2 respiration by 13.6-23.3% for soils sampled from all but the most P-rich site (ER, BR, and EV were responsive to P amendment, while BV, which also had the greatest total and available P, was not) (Table 2-1; Figure 2-1). This is consistent with previously measured soil carbon flux response following P fertilization (Cleveland and Townsend 2006) and a 20% and 16% increase in litter and soil respiration, respectively, following 158 days of P fertilization in French Guiana tropical soil (Fanin et al., 2012). While the conditions in our microcosms did not simulate closely *in-situ* environmental conditions, these values represent reference points for assessing increased respiration in field bulk soil after P fertilization.

Metatranscriptomics revealed that P amendment of low-P ridge soils (i.e., ER and BR) increased the relative transcription of numerous genes involved in cell growth and replication, including genes involved in cell division and the biosynthesis of peptidoglycan, glycerophospholipids, and purines and pyrimidines (DNA, RNA, ATP, etc. precursors) (Figure 2-3) for. A population-level investigation of cell division and cell wall (*dcw*) operon gene expression indicated that P limitation universally constrained microbial cell proliferation, at least for the most dominant taxa whose genome sequences could be recovered (Figure 2-4).

In addition to the acute growth response with lab-scale P amendment, there were positive correlations between field soil measurements of available P and MBP, MBC or MBN (Figure A-3), reflecting greater sustained biomass with greater long-term *in-situ* soil P. These results are consistent with previous studies showing an increase in microbial biomass with long-term P fertilization at tropical field locations (Liu et al., 2012; Turner and Wright 2014). In the Turner and Wright 2014 study, the authors suggested that increased microbial biomass could be due to increased plant litter inputs resulting from P fertilization, but remarked that at an adjacent site, in which soil was supplemented with doubled plant litter inputs, no increase in microbial biomass was observed (Sayer et al., 2012). Thus, while the amount and quality of organic carbon is traditionally thought to constrain soil microbial biomass and activities (Wardle, 1992; Demoling et al., 2007), our results suggest that P also directly constrains microbial cell growth and proliferation in P-limited tropical soils. This is likely due to the essentiality of P for many core cellular infrastructure components, such as DNA, RNA, ATP, NADPH, and glycerophospholipids. In order to sustain enhanced growth, microbes would also need to assimilate other nutrients and heighten catabolic functions. Accordingly, genes involved in nutrient import and/or assimilation (namely, nitrogen, sulfur, and potassium) and catabolism (*e.g.,* cytochrome biogenesis and, to a limited degree, common electron transport chain-relevant oxidoreductases) generally had greater relative transcription with P amendment (Figure 2-3; Figure A-9).

Previous research has demonstrated how conditions of excess organic carbon but inadequate P can result in the accumulation of intracellular α-glucosyl polymers in certain microbial isolates (Zevenhuizen 1966; Dephilippis et al., 1992; Woo et al., 2010). Accordingly, genes involved in the biosynthesis of α-

glucosyl polymers (*e.g.*, glycogen) had greater relative abundances in low-P ridge soils (i.e., ER and BR) relative to P-rich BV soil metagenomes (Figure 2-2). Phosphorylase genes specific to α-glucosyl polymer degradation were also more abundant in ridge soils. Soil microbiota that accumulate α-glucose polymers under P-limiting, carbon-rich conditions could benefit from metabolizing α-glucosyl substrates with phosphorylase enzymes because phosphorylase activities are impaired at low P concentrations (Derensy-Dron et al., 1999; Jaito et al., 2014). Consistent with these interpretations, we identified several genomic segments encoding functions to degrade α-glucosyl substrates (including glycogen/maltodextrin phosphorylase) accompanying genes involved in glycolysis and/or the pentose phosphate pathway (Figure 2-5). These segments could represent a streamlined pathway where stored organic substrates are utilized to generate major biosynthetic precursors and heighten cellular activities through energy-yielding catabolism upon P conditions favorable for cell proliferation and growth. Enhanced usage of stored organic substrates, as evidenced by increased relative transcription of these recovered pathways (Figure 2-5a) and all surveyed carbohydrate-active phosphorylase genes (Figure A-8), likely contributed to elevated CO_2 respiration with P amendment. Consistent with these interpretations, P amendment of low-P ridge soils increased measured α-glucosidase (responsible for the cleavage of terminal glucose residues from a variety of α-glucosyl substrates) activity by +46.0% (near-significantly; p-value<0.1), making it the most P-responsive SOC-degrading enzyme assessed in this study. Therefore, it appeared that metatranscriptomics provided, in general, consistent results with enzyme assay analysis, albeit the former offered higher resolution and sensitivity. Further, ridge soils with larger MBC:MBP exhibited greater respiration response to P amendment (Figure A-3d);

this could be attributable to increased use of intracellular, stored organic substrates in low-P ridge soil incubations that were amended with P.

Conversely, phosphorylase genes specific to β-glucosyl polymers (traditionally represented by plant-derived substrates - *e.g.*, cellulose, hemicellulose) were more abundant in P-rich BV soil metagenomes (Figure 2-2). Direct metabolism of plant-derived substrates (β-glucosyl polymers) using phosphorolysis could be metabolically advantageous under P-rich conditions because it conserves ATP that is otherwise needed to phosphorylate glucose following hydrolysis (with glucokinase). These results implicate phosphorylase genes for different substrate classes (β-glucosyl, α-glucosyl polymers) as a trait reflecting soil P bioavailability, and provides mechanistic insights into how microbial growth and activity can be coupled to, and potentially regulated, by soil P conditions.

We also observed differences in the abundance and activities of phosphorus acquisition, turnover, and storage mechanisms under high vs. low soil P conditions. For instance, the metagenome-derived abundance of *pit* (encoding low-affinity phosphate transporter) was greater in P-rich BV soils and had greater relative transcription in P-amended soils (albeit non-significantly) (Figure 2-2 and 2-3). The relative abundance of *pstS* (encoding phosphate-binding protein) was greater in low-P ridge soil metagenomes and was relatively less transcribed with P amendment. Among several phyla that differed in abundance between low-P ridge and P-rich BV soils, *Gemmatimonadetes* was 7.5-fold more abundant in BV communities (Figure A-6). While few members from *Gemmatimonadetes* have been cultivated to date, it is thus far represented mostly by organisms capable of polyphosphate accumulation under conditions of high P but otherwise unfavorable for growth (Zhang et al., 2003; DeBruyn et al.,

2013; Pascual et al., 2016). Further, there was a 2-fold greater relative abundance of *ppx* (encoding exopolyphosphatase, a polyphosphatase specific to long-chain polyphosphates) in P-rich BV metagenomes (Figure 2-2). Phosphodiesterase activity was greater than the activities of organic carbon degrading enzymes for low-P ridge soils, which was consistent with previous evaluations of tropical soil phosphoesterase activity (Turner and Wright 2014, Yao et al., 2018). Taken together, adaptation to P limitation likely includes greater reliance on inorganic phosphate binding and organic phosphorus recycling, decreased reliance on low-affinity P uptake, and lower abundance of taxa and traits involving accumulation of long-chain polyphosphates.

The relative transcription of genes involved in protein maintenance and turnover, environmental stress tolerance, and ribosome biogenesis and recycling were greater in control (not P amended) incubation soils (Figure A-9). These findings are consistent with previous studies showing high overlap in cellular protective mechanisms operating under nutrient starvation and other stress types (Spence et al., 1990; Givskov et al., 1994). Studies on microbial isolates have found that nutrient-starved cells were more resistant to oxidative stress and displayed greater thermotolerance than actively growing cells (Rockabrand et al., 1995; Petti et al., 2011), presumably due to the need to survive environmental fluctuations over larger generation times. Heightened ribosome production and recycling under P-limited conditions may reflect a continual need to divert P used for rRNA for other cellular functions (and vice-versa, from those uses back to translational machinery). Barnard and colleagues (Barnard et al., 2013) demonstrated how soil microbiota adapted to oscillating growth conditions often accumulate ribosomes during stationary growth phases, presumably to respond more rapidly upon more favorable nutrient acquisition conditions. The

aforementioned study also reported a decline in ribosome levels during conditions of elevated growth for responsive taxa. Altogether, greater relative expression of genes involved in cell upkeep, environmental stress tolerance, and core functionality may reflect suboptimal growth conditions.

The RNA-DNA integration strategy employed here, *i.e.*, normalizing population-specific transcriptional activity to *in-situ* (DNA) abundance, can be useful for comparing community metatranscriptomes where high compositional heterogeneity is an inherent environmental feature (*e.g.,* soils). This approach could also be combined with qPCR or inclusion of internal standards prior to nucleic acid extraction in order to estimate absolute abundances of target genes or genomes, if desirable. For studies enabling the recovery of several high quality metagenome-assembled genomes from complex soil communities, the transcription of certain functions could also be evaluated against baseline cellular expression/activities (*e.g.,* average transcription of all genes or core cellular functions like the *dcw* operon) or between functions where contrasting modes of response are expected (*e.g.,* high vs. low-affinity P-uptake mechanisms encoded by the same genome).

2.6. Conclusions

In this study, community composition (determined with metagenomics) and activity assessments (including CO_2 respiration, enzyme assays for organic carbon and phosphorus decomposition, and metatranscriptomics) together revealed how tropical soil microbiota respond acutely to an alleviation of P constraints. Further, functional signatures revealed community traits reflecting long-term adaptation to growth-restricting P conditions. Several of the key genes and pathways, *e.g.,* abundance of genes for α-glucosyl polymer biosynthesis,

phosphorylases specific to stored vs. plant-derived substrates, and competing modes of phosphorus acquisition, could serve as biomarkers of microbial adaptation to changes in nutrient availability. Genomic indicators of environmental change are candidates for incorporation into ecosystem models. Further, the gene and partial genome sequences reported here could provide reference data for PCR assays to monitor the abundance of specific indicator genes for nutrient status in field samples. It would also be important to test these findings against other tropical forest locations, *e.g.,* long-term fertilization studies, in order to better distinguish between the long-term and acute effects of greater P availability, and to determine how the strength and consequences of P limitation compare to other major nutrients (N, Ca, K, etc.).

2.7. Acknowledgements

This research was sponsored by the Oak Ridge National Laboratory (ORNL) Laboratory Directed Research and Development Program (to MAM); the ORNL 'GO! Student Program' (to ERJ); and the US National Science Foundation (award 1356288 to KTK). ORNL is managed by the University of Tennessee-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Support was received by US National Science Foundation for the Luquillo Critical Zone Observatory (EAR-1331841) and the Luquillo Long Term Ecological Research Program (DEB 1239764); the USDA Forest Service International Institute of Tropical Forestry and The University of Puerto Rico provided additional support. We would also like to thank Carlos Torrens of the US Forest Service for his assistance collecting the Bisley Watershed samples.

2.8. References

Bala, G., Caldeira, K., Wickett, M., Phillips, T., Lobell, D., Delire, C., Mirin, A., 2007. Combined climate and carbon-cycle effects of large-scale deforestation. Proceedings of the National Academy of Sciences 104, 6550–6555.

Barnard, R.L., Osborne, C.A., Firestone, M.K., 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. The ISME Journal 7, 2229–2241.

Beedlow, P.A., Tingey, D.T., Phillips, D.L., Hogsett, W.E., Olszyk, D.M., 2004. Rising atmospheric CO2 and carbon sequestration in forests. Frontiers in Ecology and the Environment 2, 315–322.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57, 289–300.

Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320, 1444–1449.

Brady, N., 1990. Physical properties of mineral soils. The Nature and Properties of Soils: 91–175.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Cavaleri, M.A., Reed, S.C., Smith, W.K., Wood, T.E., 2015. Urgent need for warming experiments in tropical forests. Global Change Biology 21, 2111–2121.

Cleveland, C.C., Townsend, A.R., 2006. Nutrient additions to a tropical rain forest drive substantial soil carbon dioxide losses to the atmosphere. Proceedings of the National Academy of Sciences 103, 10316–10321.

Cleveland, C.C., Townsend, A.R., Schmidt, S.K., 2002. Phosphorus limitation of microbial processes in moist tropical forests: evidence from short-term laboratory incubations and field studies. Ecosystems 5, 0680–0691.

Cleveland, C.C., Townsend, A.R., Taylor, P., Alvarez□Clare, S., Bustamante, M., Chuyong, G., Dobrowski, S.Z., Grierson, P., Harms, K.E., Houlton, B.Z., 2011. Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan□tropical analysis. Ecology Letters 14, 939–947.

Condron, L.M., Turner, B.L., Cade-Menun, B.J., 2005. Chemistry and dynamics of soil organic phosphorus. Phosphorus: Agriculture and the Environment 87–121.

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485.

DeBruyn, J.M., Fawaz, M.N., Peacock, A.D., Dunlap, J.R., Nixon, L.T., Cooper, K.E., Radosevich, M., 2013. Gemmatirosa kalamazoonesis gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum Gemmatimonadetes. The Journal of General and Applied Microbiology 59, 305–312.

Demoling, F., Figueroa, D., Bååth, E., 2007. Comparison of factors limiting bacterial growth in different soils. Soil Biology and Biochemistry 39, 2485–2495.

Dephilippis, R., Sili, C., Vincenzini, M., 1992. Glycogen and poly-β-hydroxybutyrate synthesis in Spirulina maxima. Journal of General Microbiology 138, 1623–1628.

Derensy□Dron, D., Krzewinski, F., Brassart, C., Bouquelet, S., 1999. β□1, 3□ Galactosyl□N□acetylhexosamine phosphorylase from Bifidobacterium bifidum DSM 20082: characterization, partial purification and relation to mucin degradation. Biotechnology and Applied Biochemistry 29, 3–10.

Drebot, M., Barnes, C., Singer, R., Johnston, G., 1990. Genetic assessment of stationary phase for cells of the yeast Saccharomyces cerevisiae. Journal of Bacteriology 172, 3584–3589.

Fanin, N., Barantal, S., Fromin, N., Schimann, H., Schevin, P., Hättenschwiler, S., 2012. Distinct Microbial Limitations in Litter and Underlying Soil Revealed by Carbon and Nutrient Fertilization in a Tropical Rainforest. PLoS ONE 7, e49990.

Givskov, M., Eberl, L., Molin, S., 1994. Responses to nutrient starvation in Pseudomonas putida KT2442: two-dimensional electrophoretic analysis of starvation- and stress-induced proteins. Journal of Bacteriology 176, 4816–4824.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59, 307–321.

Harrison, A.F., 1987. Soil organic phosphorus: a review of world literature (Report). Commonwealth Agricultural Bureaux International.

Harrison, K.J., Crécy-Lagard, V. de, Zallot, R., 2017. Gene Graphics: a genomic neighborhood data visualization web application. Bioinformatics 34, 1406-1408.

Holm, J.A., Kueppers, L.M., Chambers, J.Q., 2017. Novel tropical forests: response to global change. New Phytologist 213, 988–992.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

Jaito, N., Saburi, W., Odaka, R., Kido, Y., Hamura, K., Nishimoto, M., Kitaoka, M., Matsui, H., Mori, H., 2014. Characterization of a thermophilic 4-O-β-d-mannosyl-d-glucose phosphorylase from Rhodothermus marinus. Bioscience, Biotechnology, and Biochemistry 78, 263–270.

Johnston, E.R., Rodriguez-R, L.M., Luo, C., Yuan, M.M., Wu, L., He, Z., Schuur, E.A., Luo, Y., Tiedje, J.M., Zhou, J., 2016. Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. Frontiers in Microbiology 7.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772–780.

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research 44, W242–W245.

Lillie, S.H., Pringle, J.R., 1980. Reserve carbohydrate metabolism in Saccharomyces cerevisiae: responses to nutrient limitation. Journal of Bacteriology 143, 1384–1394.

Liptzin, D., Silver, W.L., 2009. Effects of carbon additions on iron reduction and phosphorus availability in a humid tropical forest soil. Soil Biology and Biochemistry 41, 1696–1702.

Liu, L., Gundersen, P., Zhang, T., Mo, J., 2012. Effects of phosphorus addition on soil microbial biomass and community composition in three forest types in tropical China. Soil Biology and Biochemistry 44, 31–38.

Mage, S.M., Porder, S., 2013. Parent Material and Topography Determine Soil Phosphorus Status in the Luquillo Mountains of Puerto Rico. Ecosystems 16, 284–294.

McGroddy, M., Silver, W.L., 2000. Variations in Belowground Carbon Storage and Soil CO2 Flux Rates along a Wet Tropical Climate Gradient1. BIOTROPICA 32, 614.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17, 132.

Oren, R., Ellsworth, D.S., Johnsen, K.H., Phillips, N., 2001. Soil fertility limits carbon sequestration by forest ecosystems in a CO2-enriched atmosphere. Nature 411, 469.

Pascual, J., Garcia-Lopez, M., Bills, G.F., Genilloud, O., 2016. Longimicrobium terrae gen. nov., sp nov., an oligotrophic bacterium of the under-represented phylum Gemmatimonadetes isolated through a system of miniaturized diffusion chambers. International Journal of Systematic and Evolutionary Microbiology 66, 1976–1985.

Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428.

Petti, A.A., Crutchfield, C.A., Rabinowitz, J.D., Botstein, D., 2011. Survival of starving yeast is correlated with oxidative stress response and nonrespiratory mitochondrial function. Proceedings of the National Academy of Sciences of the United States of America 108, E1089–E1098.

Ramirez, K.S., Craine, J.M., Fierer, N., 2012. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. Global Change Biology 18, 1918–1927.

Rho, M., Tang, H., Ye, Y., 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Research 38, e191–e191.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rockabrand, D., Arthur, T., Korinek, G., Livers, K., Blum, P., 1995. An essential role for the Escherichia coli DnaK protein in starvation-induced

thermotolerance, H2O2 resistance, and reductive division. Journal of Bacteriology 177, 3695–3703.

Rodriguez-R, L.M., Gunturu, S., Tiedje, J.M., Cole, J.R., Konstantinidis, K.T., 2018. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. MSystems 3, e00039-18.

Rodriguez-R, L.M., Konstantinidis, K.T., 2014. Estimating coverage in metagenomic data sets and why it matters. The ISME Journal 8, 2349–2351.

Sayer, E.J., Wright, S.J., Tanner, E.V., Yavitt, J.B., Harms, K.E., Powers, J.S., Kaspari, M., Garcia, M.N., Turner, B.L., 2012. Variable responses of lowland tropical forest nutrient status to fertilization and litter manipulation. Ecosystems 15, 387–400.

Scatena, F.N., 1989. An Introduction to the Physiography and History of the Bisley Experimental Watersheds in the Luquillo Mountains of Puerto Rico (No. SO-GTR-72). U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station, New Orleans, LA.

Silver, W.L., Lugo, A.E., Keller, M., 1999. Soil oxygen availability and biogeochemistry along rainfall and topographic gradients in upland wet tropical forest soils. Biogeochemistry 44, 301–328.

Sinsabaugh, R.L., Hill, B.H., Shah, J.J.F., 2009. Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. Nature 462, 795.

Southeast Regional Climate Center, n.d. Historical Climate Data For Puerto Rico. Data retrieved November 27, 2017.

Spence, J., Cegielska, A., Georgopoulos, C., 1990. Role of Escherichia coli heat-shock proteins DnaK and HtpG (C62.5) in response to nutritional deprivation. Journal of Bacteriology 172, 7157–7166.

Tanner, E., Vitousek, P., and Cuevas, E., 1998. Experimental investigation of nutrient limitation of forest growth on wet tropical mountains. Ecology 79, 10–22.

The UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Research 43, D204–D212.

Turner, B.L., 2008. Resource partitioning for soil phosphorus: a hypothesis. Journal of Ecology 96, 698–702.

Turner, B.L., Wright, S.J., 2014. The response of microbial biomass and hydrolytic enzymes to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. Biogeochemistry 117, 115–130.

Vitousek, P.M., 1984. Litterfall, nutrient cycling, and nutrient limitation in tropical forests. Ecology 65, 285–298.

Wardle, D., 1992. A comparative assessment of factors which influence microbial biomass carbon and nitrogen levels in soil. Biological Reviews 67, 321–358.

Woo, H.M., Noack, S., Seibold, G.M., Willbold, S., Eikmanns, B.J., Bott, M., 2010. Link between Phosphate Starvation and Glycogen Metabolism in Corynebacterium glutamicum, Revealed by Metabolomics. Applied and Environmental Microbiology 76, 6910–6919.

Wood, T.E., Matthews, D., Vandecar, K., Lawrence, D., 2016. Short-term variability in labile soil phosphorus is positively related to soil moisture in a humid tropical forest in Puerto Rico. Biogeochemistry 127, 35–43.

Wright, S.J., Yavitt, J.B., Wurzburger, N., Turner, B.L., Tanner, E.V., Sayer, E.J., Santiago, L.S., Kaspari, M., Hedin, L.O., Harms, K.E., 2011. Potassium, phosphorus, or nitrogen limit root allocation, tree growth, or litter production in a lowland tropical forest. Ecology 92, 1616–1625.

Yao, Q., Li, Z., Song, Y., Wright, S.J., Guo, X., Tringe, S.G., Tfaily, M.M., Paša-Tolić, L., Hazen, T.C., Turner, B.L., Mayes, M.A., Pan, C., 2018. Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. Nature Ecology & Evolution 2, 499-509.

Zevenhuizen, L., 1966. Formation and function of the glycogen-like polysaccharide of Arthrobacter. Antonie Van Leeuwenhoek 32, 356–372.

Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., Nakamura, K., 2003. Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. International Journal of Systematic and Evolutionary Microbiology 53, 1155–1163.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.

Zhang, X., Johnston, E.R., Li, L., Konstantinidis, K.T., Han, X., 2017. Experimental warming reveals positive feedbacks to climate change in the Eurasian Steppe. The ISME Journal 11, 885–895.

CHAPTER 3: EXPERIMENTAL WARMING REVEALS POSITIVE FEEDBACKS TO CLIMATE CHANGE IN THE EURASIAN STEPPE

Ximei Zhang[#], Eric R Johnston[#], Linghao Li, Konstantinos T Konstantinidis, & Xingguo Han

[#]These authors contributed equally to this article.

Originally published on December 20th, 2016 in The ISME Journal 11, 885-895

DOI: 10.1038/ismej.2016.180

3.1. Abstract

Identifying soil microbial feedbacks to increasing temperatures and moisture alterations is critical for predicting how terrestrial ecosystems will respond to climate change. We performed a five-year field experiment manipulating warming, watering and their combination in a semiarid temperate steppe in northern China. Warming stimulated the abundance of genes responsible for degrading recalcitrant soil organic matter (SOM) and reduced SOM content by 13%. Watering and warming plus watering also increased the abundance of recalcitrant SOM catabolism pathways, but concurrently promoted plant growth and increased labile SOM content, which somewhat offset SOM loss. The treatments also increased microbial biomass, community complexity, and metabolic potential for nitrogen and sulfur assimilation. Both microbial and plant community composition shifted with the treatment conditions, and the

sample-to-sample compositional variations of both groups (pairwise β-diversity distances) were significantly correlated. In particular, microbial community composition was substantially correlated with the dominant plant species (~0.54 Spearman correlation coefficient), much more than with measured soil indices, affirming a tight coupling between both biological communities. Collectively, our study revealed the direction and underlying mechanisms of microbial feedbacks to warming and suggested that semiarid regions of northern steppes could act as a net carbon source under increased temperatures, unless precipitation increases concurrently.

3.2. Introduction

Human activities have raised the concentration of atmospheric greenhouse gases, causing a 0.85 ºC increase in the global mean temperature since 1880 (IPCC, 2014). Climate warming has profound influences on biodiversity and ecosystem functioning, which in turn can result in feedbacks to climate warming (Luo *et al.*, 2014; Steinauer *et al.*, 2015). Feedbacks of soil ecosystems to climate change remain poorly understood, despite containing more carbon (C) in the form of soil organic matter (SOM) than both aboveground plant communities and atmosphere pools combined (Grosse *et al.*, 2011; Scharlemann *et al.*, 2014). Climate change could enhance microbial decomposition of SOM, primarily by promoting enzyme activities, and lead to positive feedback through the direct release of CO_2 and/or CH_4 to the atmosphere (Heimann & Reichstein, 2008; Mackelprang *et al.*, 2011; Karhu *et al.*, 2014; McCalley *et al.*, 2014). Alternatively, soil microbial communities could possibly acclimate/adapt to climate warming, and thus have no or even negative feedback (Zhou *et al.*, 2012). A net increase or decrease in terrestrial ecosystem

carbon storage will ultimately depend on the balance between plant primary productivity, microbial C fixation, soil heterotrophic respiration, and soil C stabilization over time, and is likely to vary between ecosystem types. However, it has been challenging to identify the direction of microbial feedbacks to climate change, primarily because the recalcitrant components of SOM and the microbial groups responsible for their degradation are diverse, and SOM decomposition rates are slow (Curtis *et al.*, 2002; Handelsman *et al.*, 2007).

Temperature and precipitation are perhaps the most important factors in determining biodiversity and ecosystem functioning in terrestrial ecosystems, and precipitation patterns are expected to shift with climate warming (Dore, 2005; Groisman *et al.*, 2005). While the feedback of soil microbial communities to climate warming was found to be positive in a relatively arid prairie ecosystem (~350mm mean annual precipitation), a negative feedback was revealed in a comparatively moist prairie ecosystem (~900mm mean annual precipitation) (Nie *et al.*, 2012; Zhou *et al.*, 2012). Therefore, it is important to investigate microbial feedbacks in additional ecosystem types and under concurrent changes of precipitation and temperature. Changes in moisture and temperature conditions also affect aboveground plant communities (Yang *et al.,* 2011), which govern the type and abundance of many organic substrates provided to soil microbial heterotrophs. Thus, changes in climate should also impact microbial communities indirectly through shifts in plant composition and productivity.

The semiarid steppe ecosystem in northern China is an important component of the Eurasian grassland biome (Christensen *et al.*, 2004; Niu *et al.*, 2008), where water is a key-limiting factor and soil temperature is low from later autumn to early spring, limiting microbial activity (Liu *et al.*, 2009; Zhang *et al.*, 2014). According to climate history, air temperatures adjacent to our study site

have risen by ~2.4ºC in the past several decades (1953–2008), and summer precipitation is expected to increase in the future (Sun & Ding, 2010). Here, we conducted a long-term field experiment involving watering (W) and warming (T), mimicking increased precipitation and elevated temperature, respectively, and their combination (WT). We employed shotgun metagenomic sequencing, which can reveal the presence and relative abundance of genes responsible for SOM transformations *in-situ* and at the whole-community level, of soil microbial communities sampled after 5-years of treatment and integrated the resulting sequence data with plant community properties and soil physicochemical characteristics in order to identify the direction and mechanism of soil microbial feedbacks to climate change. The biomass and species richness of plants were also monitored in order to assess the interplay between aboveground plant and belowground microbial communities (Zhang *et al.,* 2014).

Given the dry conditions and cool autumn/winter temperatures that characterize this ecosystem, we hypothesized that watering and warming will stimulate microbial growth/biomass and also increase community diversity/complexity (hypothesis 1). Enhanced microbial activities will result in greater utilization of recalcitrant SOM substrates and soil nutrients (hypothesis 2), which will be relatable to whole-community gene abundances and the corresponding environmental indices. Furthermore, the potential synchrony in plant and microbial community assemblages under this experimental regime, attributable to existing environmental parameters, treatment conditions, plant-microbe relationships, can be assessed by comparing the sample-to-sample compositional variations of both groups (hypothesis 3).

3.3. Materials and methods

3.3.1. Experimental design in the field

This experiment started in 2005 and was part of a long-term field experiment conducted in a steppe ecosystem in the Inner Mongolia Autonomous Region of China. The study site and experimental design have been described previously (Niu *et al.*, 2008; Liu *et al.*, 2009; Yang *et al.*, 2011), and only a brief description is provided here. The site was a typical temperate zone habitat, characterized by a semiarid continental monsoon climate. Annual mean temperature was 2.1ºC, with monthly mean temperature ranging from –17.5ºC in January to 18.9ºC in July. Annual mean precipitation was ~385.5mm, with 80% occurring from June to September. Soil was chestnut soil (Chinese classification), corresponding to the Calcis-orthic Aridisol in the US Soil Taxonomy classification, with sand, silt, and clay being 62.7%, 20.3%, and 17.0%, respectively. Soil mean bulk density was ~1.31g/cm^3. This ecosystem was dominated by perennials, including *Agropyron cristatum*, *Allium bidentatum*, *Artemisia frigida*, *Cleistogenes squarrosa*, *Potentilla acaulis* and *Stipa krylovii*.

This experiment adopted a paired nested design, with watering and warming as the primary and secondary factor, respectively (Niu *et al.*, 2008; Liu *et al.*, 2009; Yang *et al.*, 2011). There were three pairs of 10m×15m plots, with one plot in each pair being randomly assigned to the watering treatment and the other to the control (see Supplementary Fig. B-1 for a graphical representation of the experimental design). In each watering plot, six sprinklers were arranged evenly in two rows, with a 5m distance between two sprinklers. Each sprinkler covered a circular area with a 3m diameter, and the six sprinklers covered the whole 10m×15m plot. In July and August, 15mm water was added weekly to the

watering plots, and thus a total amount of 120mm water (approximately 30% of annually mean precipitation) was added each year. Within each 10m×15m plot, four 3m×4m sub-plots were randomly assigned to the warming and control subplots (each with two replicates). The warming subplots were heated continuously since April 28th, 2005, using 165cm×15cm MSR-2420 infrared radiators (Kalglo Electronics, Bethlehem, PA, USA) suspended 2.5m above the ground, and resulted in a ~1.1ºC increase in soil temperature above control conditions. In the control subplot, one 'dummy' heater with the same size and shape as the infrared radiator was suspended 2.5m high, mimicking the shading effect of the heater. For the watering treatment, the sprinkler was very small and its disturbance effect on the ecosystem was negligible, so we did not install a 'dummy' sprinkler in the control. Overall, there were six replicates for each of the four treatments.

3.3.2. Measurement of soil physiochemical indexes, plant and bacterial community indexes

On August 22nd, 2010, four random soil cores (10cm deep, 3.5cm diameter) were collected from each 3m×4m sub-plot. All soil cores were taken no less than 2 inches away from a plant stalk, and the aboveground foliage of plant material was removed before sampling. The four soil cores were thoroughly mixed and passed through a 2mm sieve to remove plant roots. Part of the composited soil samples was frozen for DNA extraction, while the remaining portion was used to measure SOM content, soil total nitrogen (N) content, NH_4^+-N and NO_3^--N content, water content and pH. SOM and total N contents were quantified with the potassium dichromate-vitriol oxidization method and the Kjeldahl acid-digestion method, respectively (Bao, 2000). Soil NH_4^+-N and NO_3^--

N contents were determined on a FIAstar 5000 Analyzer (Foss Tecator, Denmark) after extraction of fresh soil with 1 mol/L KCl. Soil water content was determined as the weight loss after drying for 24hr at 105ºC. Soil pH was measured in 1:2.5 (W/V) suspensions of soil in distilled water. In addition, soil temperature at the depth of 10cm was recorded with a CR1000 datalogger (Campbell Scientific, Logan, UT, USA) at 1-h intervals from June 4th, 2005. And bacterial 16S rRNA gene abundance was quantified using real-time PCR, as described before (Zhang *et al.*, 2013a,b).

All aboveground plants were harvested in two random 0.3m×0.5m quadrats from each plot; the plants were sorted into species, and then oven-dried at 65ºC for 48h and weighted. The total weight of all these plants was calculated as the aboveground plant biomass, and the total species number was counted to represent plant richness.

For each of the soil physicochemical indexes and plant and bacterial community indexes, two-way ANOVA with block as a random factor was adopted to reveal the effect of warming and watering and their interaction. Multiple comparison was also conducted among the four treatments; in particular, the Least Significant Difference method and Tamhane's T2 method was adopted for equal and unequal variances, respectively.

3.3.3. Shotgun metagenomic sequencing, sequence annotation and taxonomic analysis

Soil DNA was extracted with the MoBio PowerLyzer PowerSoil DNA isolation kit according to the manufacturer's instructions. To obtain sufficient DNA for the shotgun metagenomic sequencing and to ensure adequate representation of soil, 4-6 replicates were conducted for each sample (0.25g soil per replicate;

1.0-1.5g of total soil used for extraction for each soil composite sample). The laboratory protocol followed the description in the Illumina Paired-End Prep kit protocol. DNA was sheared mechanically and size-selected to ~180-bp and gel purified. Sequencing was performed in the Illuminia Hiseq 2000 system at Shanghai Majorbio Bio-pharm Technology Co., Ltd for 23 samples (6 per treatment); and there was one control sample for which the library preparation and sequencing methodologies were slightly different from the others, and its sequences were incomparable to other datasets.

Shotgun sequencing resulted in 12.0±4.5 million sequences (mean±one standard deviation), or 2.4±0.9 Gbp of sequencing effort, per sample (average of 14 Gbp per treatment), and 8.6±2.3 million sequences passed quality control for downstream analysis (Supplementary Table B-1). The sequences were quality trimmed and annotated (see the details in the *Supplementary materials and methods*), and further processed with the DESeq2 package (Love *et al.*, 2014), in which the differentially-abundant metabolic pathways between the treatments were identified and the false discovery rate from multiple testing was accounted for using Benjamini–Hochberg correction (adjusted *P*-values; Benjamini and Hochberg, 1995). We specifically focused on the effect of each treatment on the abundances of genes for the catabolic processes of various carbon-complexes with different decomposability, ranging from the highly recalcitrant lipids and phenolics to the more labile monosaccharides, sugar acids and sugar alcohols. More attention was also given to the genes for nitrogen/sulfate acquisition, biosynthesis, and metabolism.

The relative abundance of various prokaryotic phyla was determined by aligning the sequences to Greengenes 16S ribosomal database (DeSantis *et al.*, 2006) (see the details in the *Supplementary materials and methods*). The relative

abundance of fungi relative to prokaryotes (archaea and bacteria) was determined by aligning these reads to the Silva database (Quast *et al.*, 2013). Bacteria were found to make up ~94% of ribosomal gene fragments, and thus we further calculated bacterial taxonomic richness by sampling equal number of sequences.

3.3.4. Estimation of community complexity and functional richness

Community complexity estimations were performed on composite metagenomes, where the sequences of five replicate samples of each treatment were combined (pooling of five samples in T, W, and WT treatments to be consistent with the control group). All merged sequences $\geq$80bp were inputted to Nonpareil, which is a statistical tool that uses sequence redundancy to derive dataset complexity estimations (*i.e.,* total sequence complexity) and the amount of sequencing effort needed to achieve a desired representation of total sequence richness (Rodriguez-R & Konstantinidis, 2014).

To exclude the influence of unequal sampling, the relatively rarer functional pathways with $< 10^{-6}$ relative abundance in each sample (there were at least 10^6 reads assigned to functional genes for each sample) were removed for the calculation of pathway richness (Fierer *et al.*, 2012).

3.3.5. Statistical analysis of microbial and plant compositional variation

Bray-Curtis and weighted UniFrac distances calculated from the relative abundances of OTUs (Operational Taxonomic Unit), and Bray-Curtis distances calculated from the relative abundances of functional genes, were used to assess the compositional variation between microbial communities (Bray &

Curtis, 1957; Lozupone *et al.*, 2007). Bray-Curtis distances were also calculated from the biomass of plant species to represent plant community compositional variation of plots from which metagenomes were derived. Principal coordinate analysis (PCoA) was used to visualize the relative differences in community taxonomic and gene functional composition among different treatments (Anderson, 2003). Permutational multivariate analysis of variance (PERMANOVA) was further used to reveal the effects of experimental treatments on the taxonomic and functional composition (Anderson, 2005). Linear regression analyses were used to identify the relationships between these PCoA axes. Linear regression analyses were also used to identify the relationships between the relative change in phylum or functional gene abundance caused by different treatments.

3.3.6. Mass analysis between subsets of environmental indices and microbial community dissimilarity matrices

Mash, a tool that uses kmers to compare the sequence composition between metagenomes, was used to calculate sample-sample dissimilarities (options: -k 25 –s 100000) (Ondov *et al.*, 2015). The distance matrix obtained from Mash and from weighted UniFrac distances of 16S rRNA derived OTUs were used to test for correlation between microbial community structure and biotic and abiotic variables with “bioenv” (Clarke & Ainsworth, 1993), an analysis contained in the R package “vegan”. In short, this function calculates Euclidean distances between samples and attempts to determine the subset of environmental variables that correlates best to the user provided distance matrix. Subsets are scored and ranked with Spearman correlation. The Mash distances were also compared to plant community Bray-Curtis distances (see above) to

determine if the two distance matrices were significantly correlated using the Mantel test.

3.3.7. Accession numbers

The pyrosequencing reads were deposited in the Sequence Reads Archive database of the National Center for Biotechnology (accession no. SRA057669), and the shot-gun metagenomic data was deposited in MGRAST with the project name of “Multifactorial Environmental Changes in Inner Mongolia of China”.

3.4. Results and discussion

3.4.1. Microbial feedback direction under climate change

The warming treatment raised soil temperature at the sampling depth by 1.1ºC, on average, above the temperature of the control plots, which represented a small but significant shift in temperature relevant for climate change. Two-way ANOVA revealed that warming and watering had significant interactive effect on the SOM content ($P < 0.05$; Table 3-1). In particular, warming alone decreased the SOM content by about 13.0% relative to the control, but the decrease was smaller/negligible under watering (4.9%) and its combination with warming (3.8%; Table 3-1). Our shotgun sequencing covered about 30% of the total diversity (sequence richness) in each treatment (Supplementary Fig. B-2) based on estimates using Nonpareil (Rodriguez-R & Konstantinidis, 2014). While this coverage may appear relative low, the large number of replicated samples per treatment (5-6 replicates) and the fact that all datasets were similar in size, which

makes comparisons among datasets robust (Rodriguez-R. and Konstantinidis, 2014), offset the low coverage and provided for meaningful comparisons.

Table 3-1: The effect of experimental treatments on the biotic and abiotic indices. Lower case letters (a, b) and upper case letters (A, B) indicate $P < 0.10$ and < 0.05 in the multiple comparison, respectively.

Indices	Mean (standard error) under each treatment				*P* value from two-way ANOVA			
	Control	T	W	WT	Block	T	W	W*T
SOM content (g kg^{-1} soil)	12.22(0.68)^{a}	10.55(0.62)^{b}	11.57(0.53)^{ab}	11.68(0.64)^{ab}	0.001	0.071	0.637	0.041
Soil total N content (g kg^{-1} soil)	2.52(0.11)	2.48(0.08)	2.47(0.08)	2.36(0.08)	<0.01	0.186	0.106	0.481
Soil NH_4^+-N content (mg kg^{-1} soil)	8.46(0.33)^{A}	7.37(0.3)^{B}	8.25(0.3)^{AB}	8.31(0.29)^{A}	0.738	0.120	0.269	0.087
Soil NO_3^--N content (mg kg^{-1} soil)	5.18(0.75)^{B}	4.93(0.58)^{B}	10.74(1.75)^{A}	8.55(0.88)^{A}	0.027	0.208	<0.01	0.314
Soil pH	7.18(0.11)^{B}	7.09(0.07)^{B}	7.41(0.06)^{A}	7.35(0.03)^{AB}	0.574	0.345	0.004	0.841
Soil water content (kg kg^{-1} soil)	0.067(0.005)^{AB}	0.061(0.008)^{B}	0.110(0.028)^{A}	0.095(0.022)^{AB}	0.030	0.141	<0.01	0.459
Soil temperature (ºC)	14.6(0.1)^{B}	15.7(0.1)^{A}	14.4(0.1)^{B}	15.3(0.1)^{A}	0.513	<0.01	<0.01	0.025
Aboveground plant biomass (g m^{-2})	42.83(3.64)	42.37(9.57)	43.03(5.08)	51.39(5.29)	0.062	0.494	0.427	0.449
Plant species richness	10.17(1.22)^{AB}	8.17(1.05)^{B}	13.17(1.01)^{A}	12(0.82)^{A}	0.808	0.159	0.005	0.704
Bacterial 16S *rRNA* gene abundance (10^9 per g soil)	9.56(0.59)	11.42(1.01)	11.62(1.23)	12.02(0.86)	0.076	0.115	0.169	0.336
Bacteria relative abundance	0.934(0.005)^{B}	0.939(0.003)^{B}	0.953(0.002)^{A}	0.944(0.002)^{AB}	0.051	0.388	0.001	0.040
Fungi relative abundance	0.014(0.004)^{A}	0.007(0.002)^{B}	0.006(0.001)^{B}	0.008(0.001)^{AB}	0.786	0.298	0.071	0.028
Archaea relative abundance	0.051(0.004)^{AB}	0.054(0.004)^{A}	0.042(0.002)^{B}	0.048(0.001)^{AB}	0.018	0.080	0.006	0.580
Bacteria richness	541(15)	564(15)	606(26)	624(71)	0.161	0.589	0.111	0.961
Functional richness	1634(4)^{b}	1640(7)^{ab}	1653(4)^{a}	1656(10)^{a}	0.779	0.582	0.031	0.887

Watering, warming, and their combination consistently increased the relative abundance of catabolic pathways specific to highly recalcitrant carbon-complexes, decreased those of medium decomposability, and had inconsistent effects on those of labile carbon-complexes (*e.g.,* poly-, di- and mono-

saccharides, sugar acids, and alcohols) relative to the control samples. Specifically, genes involved in lignin degradation were 22.3% and 18.3% more abundant in W and WT treatments, respectively, compared to control ($P < 0.05$; Fig. 3-1). Observed differences in lipid catabolic pathways over control were +12.7%, +10.9%, and +18.3%, in W, T, and WT treatments, respectively ($P < 0.05$; Fig. 3-1). Genes for benzoate catabolism differed in abundance from control by +12.2%, +12.3%, and +12.1% in W, T, and WT treatments, respectively ($P < 0.10$; Fig. 3-1). Additionally, described catabolic pathways specific to phenolic or carboxylic acids generally increased in abundance over control for all three treatments in most cases (Fig. 3-1). In contrast, genes responsible for the degradation of complex carbohydrates, such as hemicellulose xylan (-13.5% in W, -14.0% in T, -11.7% in WT; $P < 0.01$) and polysaccharide starch (-33.9% in W, -41.0% in T, -29.6% in WT; $P < 0.05$; Fig. 3-1), decreased in abundance for treatment groups relative to control.

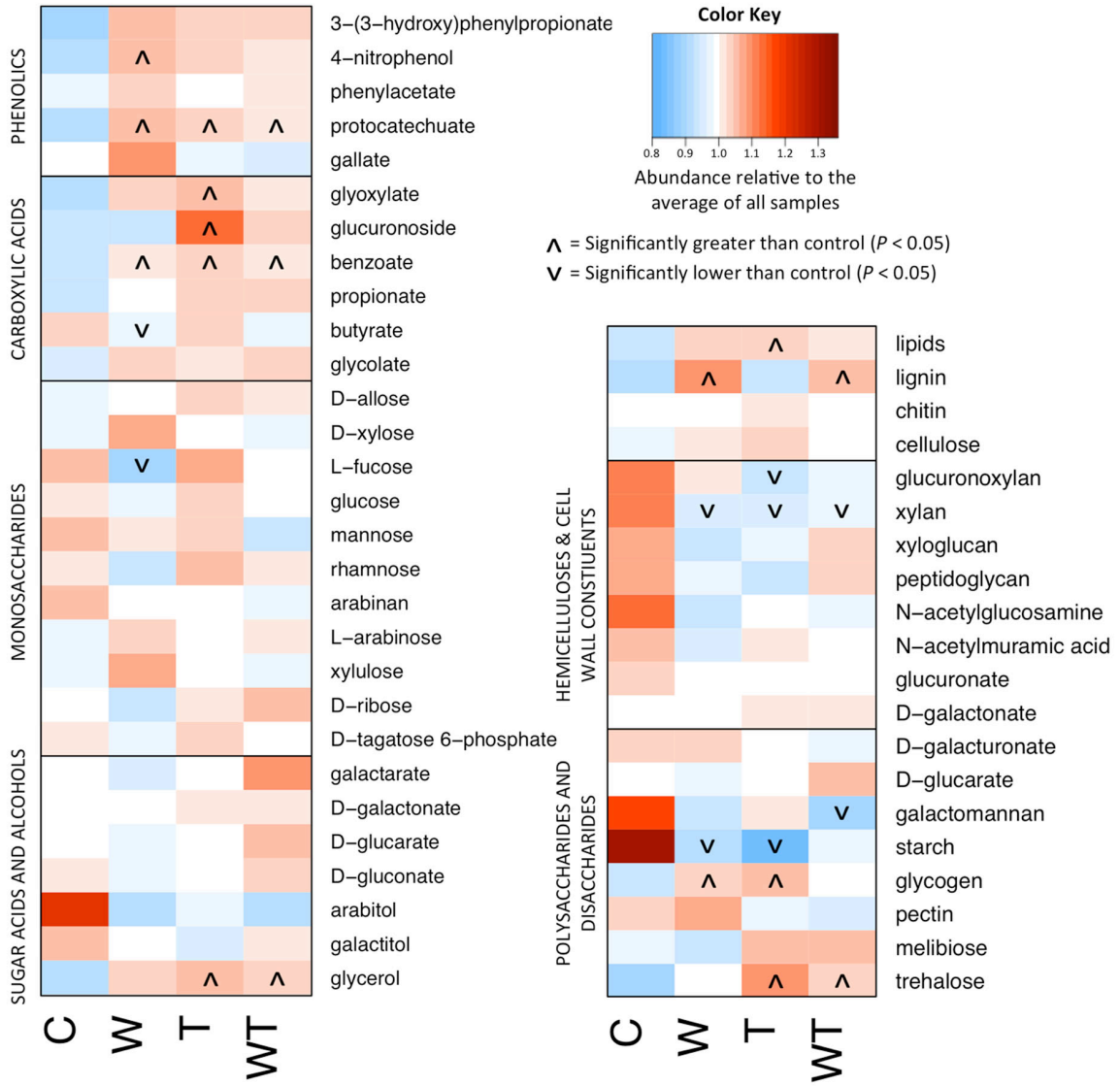


Figure 3-1: Log2-fold differences in gene abundance between treatment (W, T, WT) and control soil samples for select SOM degradation pathways. The raw count data underwent a variance-stabilizing transformation in DESeq2, which is used for logarithmically distributed count data with low mean values that tend to have high variance. Pathways denoted with an upward (increase relative to control) or downward (decrease relative to control) carat were significantly different ($P < 0.05$; see key).

A recent study found that, among all biological/physicochemical indices (e.g., pH and water content), gene abundance was the best predictor of soil element-cycling rates because it integrated the information of both environmental history and recent process activity (Petersen *et al.*, 2012). Similarly, a study on soil gene-enzyme relationships demonstrated that, when accounting for existing microbial community structure and major environmental indices, gene relative abundance was an adequate predictor of the associated enzyme activity (Trivedi *et al.*, 2016). Thus, while the increased abundance of pathways for recalcitrant SOM substrates do not represent a direct measure of enhanced microbial activity, these results suggest that microbes increased their usage of recalcitrant SOM substrates as a result of the experimental treatment conditions. This could be attributable to carbon compounds of medium decomposability becoming less abundant after five-years of treatment conditions (Fig. 3-1). Consistent with these interpretations, warming promoted microbial utilization of recalcitrant SOM such as phenols, measured by BIOLOG EcoPlates as described previously (Zhang *et al.*, 2013c) and thus decreased the total SOM content (Table 3-1), with decreases in both the labile and recalcitrant fractions of SOM, which was quantified for soil samples taken at the same time (August 2010) from this field experimental system (Song *et al.*, 2012). Meanwhile, warming had little effect on aboveground plant biomass (Table 3-1). Taken together, these results demonstrated that soil microbial communities of the Eurasian steppe under warming degraded more SOM, causing positive feedbacks to climate warming.

Experimental watering promoted an increase in plant richness ($P < 0.05$, Table 3-1); this result provided a possible explanation for the difference in SOM content observed between warming and both watering groups; *i.e.*, an elevated diversity/input of plant organic materials into the soil, as reported previously

(Song *et al.*, 2012), partly counteracted the effect of microbial degradation and caused the change in the total SOM content to be smaller/negligible (Table 3-1).

3.4.2. Intensified assimilation of nitrogen and sulfur into microbial biomass

Genes involved in the uptake and biosynthesis of nitrogen (N) and sulfur (S) compounds were generally more abundant in treatment relative to control metagenomes (Fig. 3-2). Specifically, sulfate binding (*sbp*) and import (*cysAWT)* genes were, on average, 13.5%, 9.7%, and 7.8% more abundant in W, T, and WT groups, respectively. Following S acquisition genes, pathways responsible for several steps involved in assimilatory sulfate reduction and incorporation of S into new biomass were also more abundant in treatment datasets. These genes included sulfate adenyltransferase (*cysN*, *cysD,* bifunctional *nodQ*), phosphoadenosine phosphosulfate reductase (*cysH*), and sulfite reductase (*sir*) (Fig. 3-2). While genes related to nitrate and ammonium binding and acquisition showed mixed results, assimilatory nitrate reductase gene (*nasA*) was significantly more abundant by 27.5% and 26.6% in W and WT samples, respectively, relative to control ($P < 0.05$, Fig. 3-2). Although nitrate reductase genes *napA* and *narG* were more abundant in all treatment groups, genes specific to dissimilatory nitrate reduction (*nrfA*, *nrfH*) displayed mixed results, and genes specific to denitrification (*nirK/S*, *norBC*, *nosZ*) generally decreased in abundance across all treatment groups, relative to the control group. These results suggested that watering and warming treatments stimulated the growth of soil microorganisms and favored microbes capable of acquiring S and N inorganic nutrients. Consistent with these gene abundances, microbial biomass carbon and nitrogen were found to increase under these treatments, as reported previously (Zhou *et al.*, 2013). In addition, the 16S rRNA gene abundance of soil

bacterial communities based on quantitative PCR also showed an increase trend under the three treatments (19-26% increase), although it was non-significant statistically ($P > 0.10$; Table 3-1).

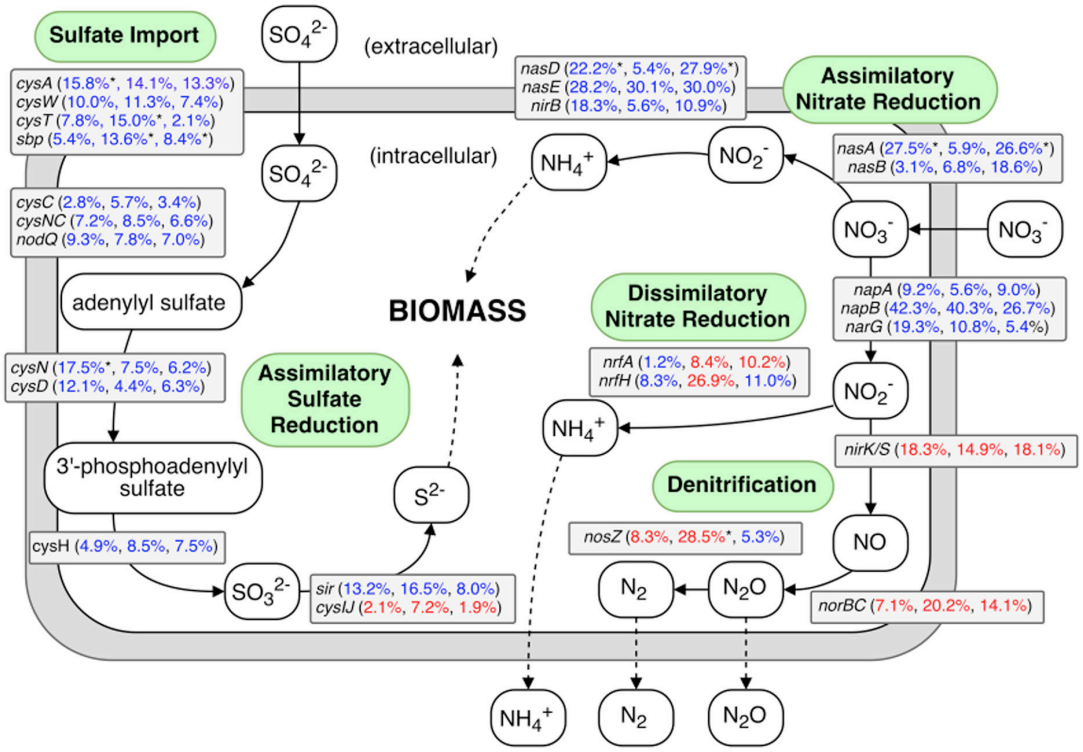


Figure 3-2: A diagram representing selected nitrogen and sulfur pathways and the difference in abundance of the underlying genes between control and treatment metagenomes. Percentages are ordered like this: watering vs. control, warming vs. control, and watering+warming vs. control. Percentages given in blue or red represent an increase or decrease in the abundance of the corresponding genes, respectively. Percentages denoted with an asterisk (*) represent abundance changes that were statistically significant ($P < 0.05$, DESeq2).

3.4.3. Stimulation of microbial community complexity and functional richness

Sequence-redundancy analysis using Nonpareil (Rodriguez-R & Konstantinidis, 2014) reveals that the composite metagenome of each treatment (W, T, and WT) had greater overall sequence diversity than the composite control metagenome (Supplementary Fig. B-2), reflecting more complex microbial communities associated with these treatments. Thus, a larger sequencing effort is required to achieve a similar representation of the corresponding communities relative to the control one. For example, to reach 50% coverage, 13.4, 18.2, 15.7 and 16.6 billion base pairs of sequencing depth would be acquired for the control, watering, warming and watering plus warming treatment, respectively. By combining five metagenomes for each sample group, approximately 30% of estimated community richness was represented in each case, with un-sampled diversity presumably belonging to sequences of comparatively rarer taxa.

Consistent with the result of community sequence complexity, there was greater microbial functional richness in the three treatment groups relative to the control, and this trend is especially obvious under watering and watering plus warming ($P < 0.05$; Table 3-1), which was congruent with the two treatments having the highest estimated community sequence complexity (see above). Both the relative abundance and species richness of bacteria was greater in the three treatments relative to the control (Table 3-1), and bacterial rRNA or functional genes made up more than 90% of the total, indicating that the stimulated community complexity and functional richness primarily resulted from bacteria rather archaea and fungi.

Warming and watering treatments likely resulted in more favorable metabolic conditions for soil microbes and thus increased the functional/metabolic diversity of their corresponding soil microbial communities. Several recent studies have demonstrated a positive correlation between gene diversity and ecosystem functioning (Salles *et al.*, 2012; Singh *et al.*, 2014; Trivedi *et al.*, 2016); thus, the elevated complexity of these soil microbial communities indicated that the overall functional potential might have increased. Thus, it is likely that an overall increase in community diversity is at least partially responsible for the enhanced degradation of recalcitrant SOM in W, T, and WT groups (Song *et al.*, 2012).

3.4.4. High correlation between microbial composition and plant and soil indices

Experimental watering resulted in significant shifts in plant community composition ($P < 0.05$). In particular, it primarily favored the growth of grasses and semi-shrubs compared to non-gramineous forbs, as reported previously (Yang *et al.*, 2011). There was a significant positive correlation between plant community variation (*i.e.,* pairwise Bray-Curtis β-diversity distances) and the compositional variation of microbial communities (using Mash distances) (Mantel test, $R^2 = 0.35$, $P < 0.01$) (see also Supplementary Fig. B-3). These results imply that microbial communities with similar local plant communities tend to be more similar than those with different plant communities. The specific ecological mechanisms governing these patterns require additional investigation.

Furthermore, correlations were obtained between microbial community sample-to-sample similarity and environmental indices and plant biomass data (Clarke & Ainsworth, 1993). For distances based on kmer-composition of the

metagenomes using Mash (Ondov *et al.*, 2015), the maximum correlation obtained was 0.544 when using the abundance of 6 of the 7 most abundant plant species (Supplementary Data S1). For distances derived from 16S rRNA OTUs (weighted unifrac), the maximum correlation achieved was 0.542, using 3 of the 7 most abundant plant species, NH_4^+-N content, NO_3^--N content, and soil temperature (environmental indices included in assessment: Total C, Total N, NH_4^+-N, NO_3^--N, soil moisture, pH, temperature, aboveground plant biomass, plant species richness). It is worth noting that distance matrices derived from sequence annotations, *i.e.,* Euclidean or Bray-Curtis distances derived from counts of reference proteins (*e.g.*, Swiss-Prot) or 16S-derived taxonomic summaries, which represent the more common practices, only provided weak correlations with plant and environmental indices (correlation coefficients <0.25). These results revealed that community composition was highly relatable to measured environmental and more so to plant parameters of the ecosystem, which is uncommon for soils due to their high complexity and heterogeneity compared to other habitats, especially when sampling bulk soil (as opposed to the rhizosphere). It also demonstrates an improvement in associating microbial community composition with abiotic and biotic indices by incorporating sequence relatedness (weighted UniFrac) or sequence composition (Mash; k-mer based) over commonly used functional gene annotation techniques. In previous studies, only a small part of microbial community variation could be explained by plant and soil indices, with the remaining fraction often attributed to under-sampling and ecological stochastic processes (Ramette & Tiedje, 2007; Ge *et al.*, 2008; Wang *et al.*, 2015).

Compared to the moderate-to-high correlation achieved between plant species community distances and Mash distances, a correlation of only 0.200

was observed when using solely abiotic environmental indices. This is likely because plant community composition is also reflective of local environmental conditions, as well as interactions between plants and belowground microbial communities, which might be un-relatable to soil conditions. This finding suggests that an alteration in environmental conditions directly affecting one biological community (e.g., plants) would influence the other group (belowground microbes) indirectly (if not directly), due to associations between biological communities. However, further investigations are required in order to obtain a more quantitative understanding of the direct and indirect associations between plant and soil microorganisms.

In agreement with the abovementioned results of the correlation analysis, PERMANOVA revealed that watering and warming significantly changed the taxonomic and functional structure of soil microbial communities, respectively ($P < 0.05$; Fig. 3-3a,b and Supplementary Fig. B-4), and the PCoA axes of the taxonomic and functional composition showed significant linear relationships to each other ($P < 0.05$; Fig. 3-3c). Consistently, PERMANOVA also revealed that watering significantly altered the taxonomic composition of soil bacterial communities measured with 454 pyrosequencing targeting at 16S rRNA genes, as reported previously (Zhang *et al.*, 2013b). Meanwhile, the treatments had similar effects on soil microbial communities because the relative changes in phyla/functional abundance caused by the three treatments were positively correlated (Supplementary Fig. B-5).

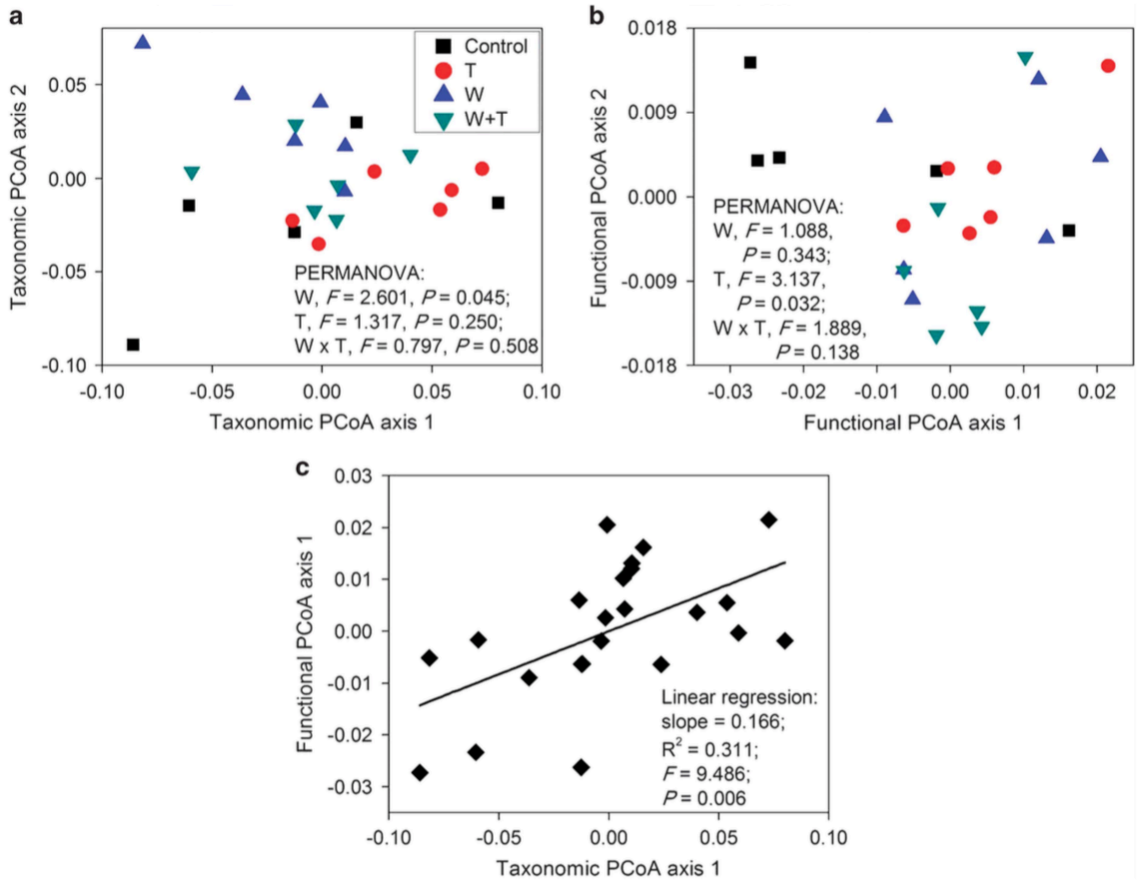


Figure 3-3: Associations between microbial taxonomic and functional community structure between treatment conditions. Effects of experimental treatments on the taxonomic (a) and functional (b) composition of soil microbial communities, and the relationship between the taxonomic and functional compositions (c). The proportion of variation explained by taxonomic PCoA axes 1 and 2 is 41.41% and 22.10%, respectively. The proportion of variation explained by functional PCoA axes 1 and 2 is 44.92% and 20.77%, respectively.

As stated above, besides plant species biomass, microbial community composition was also correlated with soil NH_4^+-N and NO_3^--N content, indicating that soil nitrogen content may have played an important role in structuring microbial community. Meanwhile, soil NH_4^+-N content decreased under these treatments (Table 3-1), whereas microbial genes for nitrogen assimilation (Fig. 3-2) and the activities of nitrogen-acquisition enzymes of microbial communities (Zhou *et al.*, 2013) increased under treatment conditions, consistent with an overall higher demand for nitrogen. These results suggest that the effect of plants on belowground microbial communities might be mediated, as least in part, by higher demand for nitrogen. Experimental methods aimed at uncovering the extent to which plants and belowground microbial communities (the biological groups primarily mediating C fixation and release from soils) respond synchronously vs. independently to these types of environmental change should continue to be a focal point in future research endeavors.

3.5. Conclusions

The SOM content is as high as ~2.4 kg/m^2 in the 0-10cm soil layer, representing ~1.2% of the total soil weight in this semiarid steppe ecosystem. Because most soil organic material is very complex and recalcitrant, total SOM content often seems to be very stable for long periods of time (Lützow *et al.*, 2006; Mikutta *et al.*, 2006), even under the influence of climate warming (Zhou *et al.*, 2012). Here, we found a significant decrease in the SOM content caused by just five years of moderate climate warming of ~1.1ºC (Table 3-1), providing strong evidence of microbial positive feedback to climate warming. This is because warming (and also watering) enhanced microbial population size (more so taxa capable of degrading recalcitrant SOM compounds and with pathways

involving the acquisition of inorganic N and S nutrients), and stimulated community diversity and complexity. Meanwhile, we find that a concurrent increase in precipitation will wholly or partially counteract soil microbial positive feedback. The close synchrony between plant and belowground communities across this experimental regime underscores the value and necessity of investigating potential co-responses and interactions between groups in future studies, even for bulk (non-rhizosphere) soil. It will also be important to assess the gene activity level (e.g., metatranscriptomics), in addition gene presence (e.g., metagenomics), and obtain more samples to test for seasonal and inter-annual dynamics in order for a more complete picture to emerge.

3.6. Acknowledgements

We thank Professors Shiqiang Wan, Quansheng Chen, Wenming Bai and many others for setting up the experiment; Yi Ren, Qiuping Hu, and many others in Shanghai Majorbio Bio-pharm Technology Co., Ltd. for help in sequencing. This research was supported by the National Key Research and Development Program (2016YFC0500702) and the Strategic Priority Research Program of CAS (XDB15010404) of China, and by the U.S. Department of Energy (award DE-SC0004601). We also thank three anonymous reviewers for helpful suggestions.

3.7. References

Anderson MJ. (2003). PCO: a FORTRAN computer program for principal coordinate analysis. Department of Statistics, University of Auckland, New Zealand.

Anderson MJ. (2005). PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. Department of Statistics, University of Auckland, New Zealand.

Bao SD. (2000). *Soil and Agricultural Chemistry Analysis*. China Agriculture Press.

Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Statist Soc Ser B-Methodol* **57:** 289–300.

Bray JR, Curtis JT. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27:** 325–349.

Christensen L, Coughenour MB, Ellis JE, Chen Z. (2004). Vulnerability of the Asian typical steppe to grazing and climatic change. *Climatic Change* **63**: 351–368.

Clarke KR, Ainsworth M. (1993). A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* **92:** 205–219.

Curtis TP, Sloan WT, Scannell JW. (2002). Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99:** 10494–10499.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Dore MHI. (2005). Climatic change and changes in global precipitation patterns: what do we know? *Environ Int* **31:** 1167–1181.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL *et al.* (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA* **109:** 21390–21395.

Ge Y, He J, Zhu Y, Zhang J, Xu Z, Zhang L *et al.* (2008). Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? *ISME J* **2:** 254–264.

Groisman PY, Knight RW, Easterling DR, Karl TR, Hegerl GC, Razuvaev VN. (2005). Trends in intense precipitation in the climate record. *J Climate* **18:** 1326–1350.

Grosse G, Harden J, Turetsky M, McGuire AD, Camill P, Tarnocai C *et al.* (2011). Vulnerability of high-latitude soil organic carbon in North America to disturbance. *J Geophys Res* **116**: G00K06.

Handelsman J, Tiedje J, Alvarez-Cohen L, Ashburner M, Cann IKO, DeLong EE *et al.* (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. The National Academies Press, Washington, DC.

Heimann M, Reichstein M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451:** 289–292.

Intergovernmental Panel on Climate Change (IPCC). (2014). *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. Cambridge University Press.

Karhu K, Auffret MD, Dungait JA, Hopkins DW, Prosser JI, Singh BK *et al.* (2014). Temperature sensitivity of soil respiration rates enhanced by microbial community response. *Nature* **513**: 81–84.

Liu W, Zhang Z, Wan S. (2009). Predominant role of water in regulating soil and microbial respiration and their responses to climate change in a semiarid grassland. *Glob Change Biol* **15:** 184–195.

Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Lozupone C, Hamady M, Kelley ST, Knight R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73:** 1576–1585.

Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K *et al.* (2014). Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. *Appl Environ Microbiol* **80:** 1777–1786.

Lützow MV, Kögel-Knabner I, Ekschmitt K, Matzner E, Guggenberger G, Marschner B *et al.* (2006). Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions – a review. *Eur J Soil Sci* **57:** 426–445.

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ *et al.* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480:** 368–371.

McCalley CK, Woodcroft BJ, Hodgkins SB, Wehr RA, Kim E, Mondav R *et al.* (2014). Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**: 478–481.

Mikutta R, Kleber M, Torn MS, Jahn R. (2006). Stabilization of soil organic matter: association with minerals or chemical recalcitrance? *Biogeochemistry* **77:** 25–56.

Nie M, Pendall E, Bell C, Gasch CK, Raut S, Tamang S *et al.* (2012). Positive climate feedbacks of soil microbial communities in a semiarid grassland. *Ecol Lett* **16**: 234–241.

Niu S, Wu M, Han Y, Xia J, Li L, Wan S. (2008). Water-mediated responses of ecosystem C fluxes to climatic change in a temperate steppe. *New Phytol* **177:** 209–219.

Ondov BD, Treangen TJ, Mallonee AB, Bergman NH, Koren S, Phillippy AM. (2015). Fast genome and metagenome distance estimation using MinHash. bioRxiv. http://dx.doi.org/10.1101/029827.

Petersen DG, Blazewicz SJ, Firestone M, Herman DJ, Turetsky M, Waldrop M. (2012). Abundance of microbial genes associated with nitrogen cycling as indices of biogeochemical process rates across a vegetation gradient in Alaska. *Environ Microbiol* **14:** 993–1008.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl Acids Res* **41(D1):** D590–D596.

Ramette A, Tiedje JM. (2007). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci USA* **104**: 2761–2766.

Rodriguez-R LM, Konstantinidis KT. (2014). Estimating coverage in metagenomic data sets and why it matters. *ISME J* **8:** 2349–2351.

Salles JF, Le Roux X, Poly F. (2012). Relating phylogenetic and functional diversity among denitrifiers and quantifying their capacity to predict community functioning. *Front Microbiol* **3:** 209.

Scharlemann JPW, Tanner EVJ, Hiederer R, Kapos V. (2014). Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Manag* **5:** 81–91.

Singh BK, Quince C, Macdonald CA, Khachane A, Thomas N, Al-Soud WA *et al.* (2014). Loss of microbial diversity in soils is coincident with reductions in some specialized functions. *Environ Microbiol* **16:** 2408–2420.

Song B, Niu S, Zhang Z, Yang H, Li L, Wan S. (2012). Light and heavy fractions of soil organic matter in response to climate warming and increased precipitation in a temperate steppe. *PLoS ONE* **7:** e33217.

Steinauer K, Tilman D, Wragg PD, Cesarz S, Cowles JM, Pritsch K *et al.* (2015). Plant diversity effects on soil microbial functions and enzymes are stronger than warming in a grassland experiment. *Ecology* **96:** 99–112.

Sun Y, Ding Y. (2010). A projection of future changes in summer precipitation and monsoon in East Asia. *Sci China Earth Sci* **53:** 284–300.

Trivedi P, Delgado-Baquerizo M, Trivedi C, Hu H, Anderson IC, Jeffries TC *et al.* (2016). Microbial regulation of the soil carbon cycle: evidence from gene–enzyme relationships. *ISME J* doi: 10.1038/ismej.2016.65.

Wang X, Van Nostrand JD, Deng Y, Lü X, Wang C, Zhou J *et al.* (2015). Scale-dependent effects of climate and geographic distance on bacterial diversity

patterns across northern China's grasslands. *FEMS Microbiol Ecol* **91:** doi: 10.1093/femsec/fiv133.

Yang H, Wu M, Liu W, Zhang Z, Zhang N, Wan S. (2011). Community structure and composition in response to climate change in a temperate steppe. *Glob Change Biol* **17:** 452–465.

Zhang N, Liu W, Yang H, Yu X, Gutknecht JLM, Zhang Z *et al.* (2013c). Soil microbial responses to warming and increased precipitation and their implications for ecosystem C cycling. *Oecologia* **173:** 1125–1142.

Zhang X, Barberán A, Zhu X, Zhang G, Han X. (2014). Water content differences have stronger effects than plant functional groups on soil bacteria in a steppe ecosystem. *PLoS ONE* **9:** e115798.

Zhang X, Liu W, Schloter M, Zhang G, Chen Q, Huang J *et al.* (2013a). Response of the abundance of key soil microbial nitrogen-cycling genes to multi-factorial global changes. *PLoS ONE* **8:** e76500.

Zhang X, Zhang G, Chen Q, Han X. (2013b). Soil bacterial communities respond to climate changes in a temperate steppe. *PLoS ONE* **8:** e78616.

Zhou J, Xue K, Xie J, Deng Y, Wu L, Cheng X *et al.* (2012). Microbial mediation of carbon cycle feedbacks to climate warming. *Nature Clim Change* **2**: 106–110.

Zhou X, Chen C, Wang Y, Xu Z, Han H, Li L *et al.* (2013). Warming and increased precipitation have differential effects on soil extracellular enzyme activities in a temperate grassland. *Sci Total Environ* **444:** 552–558.

CHAPTER 4: METAGENOMICS REVEALS PERVASIVE BACTERIAL POPULATIONS AND REDUCED COMMUNITY DIVERSITY ACROSS THE ALASKA TUNDRA ECOSYSTEM

Eric R Johnston, Luis M Rodriguez-R, Chengwei Luo, Mengting M Yuan, Liyou Wu, Zhili He, Edward AG Schuur, Yiqi Luo, James M Tiedje, Jizhong Zhou, & Konstantinos T Konstantinidis

Originally published on April 24th, 2016 in Frontiers in Microbiology 2016, 7:579

DOI: 10.3389/fmicb.2016.00579

4.1. Abstract

How soil microbial communities contrast with respect to taxonomic and functional composition within and between ecosystems remains an unresolved question that is central to predicting how global anthropogenic change will affect soil functioning and services. In particular, it remains unclear how small-scale observations of soil communities based on the typical volume sampled (1-2 grams) are generalizable to ecosystem-scale responses and processes. This is especially relevant for remote, northern latitude soils, which are challenging to sample and are also thought to be more vulnerable to climate change compared to temperate soils. Here, we employed well-replicated shotgun metagenome and 16S rRNA gene amplicon sequencing to characterize community composition and metabolic potential in Alaskan tundra soils, combining our own datasets with

those publically available from distant tundra and temperate grassland and agriculture habitats. We found that the abundance of many taxa and metabolic functions differed substantially between tundra soil metagenomes relative to those from temperate soils, and that a high degree of OTU-sharing exists between tundra locations. Tundra soils were an order of magnitude less complex than their temperate counterparts, allowing for near-complete coverage of microbial community richness (~92% breadth) by sequencing, and the recovery of twenty-seven high-quality, almost complete (>80% completeness) population bins. These population bins, collectively, made up to ~10% of the metagenomic datasets, and represented diverse taxonomic groups and metabolic lifestyles tuned toward sulfur cycling, hydrogen metabolism, methanotrophy, and organic matter oxidation. Several population bins, including members of *Acidobacteria*, *Actinobacteria*, and *Proteobacteria*, were also present in geographically distant (~100-530 km apart) tundra habitats (full genome representation and up to 99.6% genome-derived average nucleotide identity). Collectively, our results revealed that Alaska tundra microbial communities are less diverse and more homogenous across spatial scales than previously anticipated, and provided DNA sequences of abundant populations and genes that would be relevant for future studies of the effects of environmental change on tundra ecosystems.

4.2. Introduction

Terrestrial soil systems are residence to some of the most functionally and taxonomically diverse microbial communities known (Torsvik et al. 1990, Whitman et al. 1998, Curtis et al. 2002, Handelsman et al. 2007). An increasing amount of attention has been directed towards these communities due to human dependence on soil productivity for food and fiber, the ecosystem services they

provide (*e.g.,* water quality, nutrient cycling), and their role in producing and consuming greenhouse gases. Soil systems are estimated to contain more carbon than aboveground plant biomass and atmospheric pools combined in the form of degradable soil organic matter (or SOM) (Grosse et al. 2011). Higher land temperatures are expected to cause the release of considerable amounts of CO_2 and CH_4 to the atmosphere (Heimann and Reichstein 2008, Mackelprang et al. 2011, McCalley et al. 2014), primarily through the microbially-mediated degradation of SOM. Thus, there is an imminent need to further understand the role of soil microbes in the cycling of SOM C and other major elements, both to improve climate change predictions and possibly to mitigate climate change impacts through changes in land management practices. Tundra SOM is particularly sensitive to climate change (Jorgenson et al. 2010; Grosse et al. 2011) because low temperatures and saturated soil conditions protect organic C from microbial decomposition (McGuire et al. 2010; Lee et al. 2012; Pries et al. 2012). Furthermore, more than 50% of global soil organic C is stored in northern tundra permafrost, which only accounts for approximately 16% of the global soil area (Tarnocai et al. 2009). It is projected that permafrost may recede by 30-70% towards the end of the 21st century due to increasing temperatures (Schuur and Abbott 2011; Lawrence et al. 2012), likely resulting in enormous terrestrial ecosystem C loss.

Our ability to predict soil ecosystem functioning and resilience and to manipulate terrestrial soils for enhanced C sequestration is hindered, at least partially, by the enormous diversity and as yet uncultivated status of soil microorganisms (Handelsman et al. 2007; Whitman et al 1998). Several recent studies have employed ‘omics methodologies (*i.e.,* metagenomics, metatranscriptomics, metaproteomics, etc.) to characterize microbes and their

metabolisms present in tundra locations and have successfully assembled novel population bins (*i.e.,* consensus genome assembly from a natural population), representing organisms relevant to CH_4 and CO_2 release, a feat that opens up new opportunities to directly study the *in-situ* response of specific organisms (Mondav et al. 2014; Hultman et al. 2015). However, much remains unknown about what prokaryotic taxa dominate tundra, how much they vary in abundance across distant sites with similar environmental features, what pathways they encode and perhaps more importantly, what abiotic and biotic factors control the activity of these pathways and how environmental changes will affect that activity. It is also unclear how the genetic information present in the small volume of soil typically sampled (1-2 grams) by these previous surveys relates to ecosystem-scale responses and processes. For this, surveys that analyze multiple replicated samples are needed.

Our team has been performing warming manipulations that raised *in-situ* temperatures by 2-5ºC, simulating the effect of future climate change, for active layer soil atop permafrost at the Carbon in Permafrost Experimental Heating Research (CiPEHR) site (Alaska, USA; "AK site") (Zhou et al. 2012, Natali et al. 2011, 2014). In total, 11 soils from the CiPEHR, AK site were collected from 15-25 cm depths in 2010, after about 1.5-year of experimental warming. Only minor differences were observed between warming and control plots at the DNA level (metagenomics) for these samples (Xue et al. 2016), presumably due to the slow growth kinetics of tundra microbes. Here, we took advantage of the well-replicated sequence datasets available, and pooled them together in order to robustly address the following objectives: 1) evaluate the biogeography of microorganisms in tundra soils at the 16S rRNA gene level as well as at the individual population (whole-genome) level. The latter is a better proxy for

species since it circumvents the limitations of 16S rRNA gene related to high sequence conservation and represents an important and highly resolved unit of microbial communities (Caro-Quintero and Konstantinidis 2012). 2) Identify the similarities in taxonomic and functional gene composition in active-layer soil sampled from various Alaskan tundra locations, using soil communities from temperate locations for comparison. And, 3) assess how these tundra microbial populations might respond to major environmental perturbations such as fire events. Our work identified several highly abundant (>1% of total community) populations that are ubiquitous across the tundra ecosystem in Alaska and thus, represent important members of the indigenous communities. It also revealed that these populations are highly dynamic, and can undergo rapid genomic alternations in gene content upon major environmental perturbations.

4.3. Materials and Methods

4.3.1. Site description and sampling

The CiPEHR site was established in September 2008 at a moist acidic tundra area in Interior Alaska near Denali National Park in the Eight Mile Lake region (63º52’59’’N, 149º13’32’’W). The experimental plots were located in the discontinuous permafrost region where permafrost thaw has been observed in the past several decades. Experimental design and site description were described in detail previously (Natali et al. 2011). Generally, three experimental blocks were located approximately 100 m away from each other. In each block, two snow fences were constructed about 5 m apart in the winter. The winter warming treatment plots were located 5 m back from the leeward side of the

snow fences, while the paired control plots were at the windward side of snow fences. Soil temperature was increased in the winter warming plots due to thicker snow cover on soil surface and lower wind strength. The snow fences were removed in the spring before the snow melting to uniform hydraulic condition in both winter warming and control treatments. From 1976 to 2009, mean monthly temperature in the field ranged from –16°C in December to 15°C in July, with an annual mean temperature of –1.0°C. The average annual precipitation was 378 mm. Only C_3 plant species were observed in this area. Dominant species include *Eriophorum vaginatum*, *Vaccinium uliginosum*, some other vascular species, nonvascular feather moss and lichen. In the experimental plots, the upper 0.45-0.65 m soil was rich in organic C materials and below was mineral soil with a mixture of glacial till and windblown loess. The active layer depth was about 50 cm.

Eleven soil cores, five from control plots and six from warmed plots (there were originally twelve soil samples - one control soil was discarded as it was deemed contaminated with Oklahoma soil DNA, and thus, incomparable to other soil samples), were taken using electric drills in destructive sampling plots at the six snow fences in the beginning of 2010 growing season (May), one and half year after the initial of winter warming treatment. Soil fractions of 15-25 cm from ground surface were used for this study.

The experimental warming site at Oklahoma is located at the Kessler Farm Field Laboratory (KFFL, OK site) at the Great Plain Apiaries in McClain County, Oklahoma, USA (34°58'54"N, 97°31'14"W). Block design was applied in this field experiment and adjacent blocks were two or five meters apart. The six plots in each of the four blocks represent control, warming, half or double precipitation treatments and the combination manipulation of these treatments.

Additionally, every southern half subplot was clipped twice a year to create a coupled clipping effect. Only warming and control treatments without water and clipping effects were included in this study. Beginning in early 2009, the soil temperature in the warming treatment plots was increased by the Kalglo MRM-1215 120V, 1500W, 65 inch-long electric infrared radiators (Kalglo Electronics, Bethlehem, PA, USA) fixed at 1.5 m above the ground surface at the center of each plot. In control plots, wood "dummy" heaters were used to simulate shading effect in warming plots. The herbivores were excluded at this site to prevent grazing. The plant community undergoes a relatively rapid secondary succession in this site and new species occur every year, causing gradually change in plant community structure. Although both C_3 and C_4 plants were observed, C_3 species have dominated in recent years. Plant biomass peaks twice in late spring and early autumn every year. C_3 grass *Bromus arvensis* and C_3 forb *Vicia sativa* dominated the site in April 2010, while C_3 forb *Ambrosia trifida*, *Solanum carolinense* and C_4 grass *Tridens flavus* prevailed in August 2010. Based on Oklahoma Climatological Survey from 1948 to 1999, the mean annual temperature in this site was 16.3ºC, with the lowest monthly mean of 3.3ºC in January and the highest of 28.1ºC in July. The precipitation was unevenly distributed annually, which peaked in May and June (240 mm) and reached a low in January and February (82 mm) with an annual mean of 967 mm (Zhou et al. 2012). Soils from the layer 0-15 cm in four warming plots and four control plots were sampled in the OK, USA site using a standard soil core (2.5 cm in diameter) in Oct 2010. All samples were transported to the laboratory and stored at -80°C immediately until analyses. Any observable plant root materials were picked out before the soil was processed. Fungal community composition of CiPEHR and KFFL soil communities has been addressed previously (Penton et al. 2013).

Environmental indices for KFFL and CiPEHR, and the associated methods, are provided in Xu et al. 2016; a summary of select soil measurements (including C%, N%, SOM fractions, pH, and bulk density) is given in Table C-1.

4.3.2. DNA extraction of soil microbial community

Soil DNA was extracted using a PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) according to manufacturer's protocol. DNA quality was assessed based on spectrometry absorbance at wavelengths of 230 nm, 260 nm and 280 nm detected by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., now NanoDrop Products by Thermo Fisher Scientific). The absorbance ratios of 260/280 nm were around 1.8, and of 260/230 nm were larger than 1.7. Finally, DNA was quantified by Pico Green using a FLUOstar OPTIMA fluorescence plate reader (BMG LabTech, Jena, Germany) and used for gene array labeling and sequencing library preparation.

4.3.3. Illumina MiSeq sequencing protocol

The 16S rRNA library was prepared using methods introduced by Caporaso et al. 2011 and Caporaso et al. 2012. In brief, extracted DNA samples were diluted to 2.5 ng/μL for PCR amplification. The primer sets used to amplify the V4 region of 16S rRNA genes were constructed to adapt the barcode Illumina MiSeq (Caporaso et al. 2011): the forward PCR primer contains an Illumina adapter sequence, followed by a forward primer pad, a forward primer link and then the 515 forward primer; besides above elements for the reverse primer (806 reverse primer was used), the reverse PCR primer also contains a sample-

unique barcode sequence inserted between the reverse Illumina adapter and the reverse primer pad sequences for parallel sequencing of a sample set. The 25 μL PCR reaction system and condition was as documented in Caporaso et al. 2011. Only one PCR reaction was performed per sample. After the amplification the products were quantified using PicoGreen on a FLUOstar OPTIMA fluorescence plate reader (BMG LabTech). 100 ng PCR products from each sample were combined into one tube, ran on a 1% agarose gel at 100V for 45 min, and purified through QIAquick Gel Extraction Kit (Qiagen) column. The purified sample was quantified again using PicoGreen by triplicates, ensured the accuracy of library concentration. Then the pooled sample was diluted to 2 nM. 10 μL of 0.1 N NaOH was then added into 10 μL sample DNA for denaturation. Then the denatured DNA was diluted to 6 pM and mixed with equal volume of 6 pM Phix library in order to increase sequence diversity. Finally, the mixture (600 μL) was loaded into the reagent cartridge and run on MiSeq (Illumina, Inc., San Diego, CA, USA) for two ends by 150 bp reactions (Illumina) following manufacturer's instructions.

4.3.4. Metagenomic shotgun sequencing protocol

DNA integrity was confirmed by gel electrophoresis and sent to Los Alamos National Laboratory to run on Illumina HiSeq platform. The DNA was fragmented and the library was prepared using TruSeq Kit (Illumina) according to manufacturer's protocol. Each of the nineteen samples was sequenced in one flow cell lane with 2×150 bp paired-end format.

4.3.5. Paired-end sequence merging and quality trimming

Reads were merged using PEAR (Zhang et al. 2014) (options: -p 0.001). Both merged and unmerged reads underwent quality trimming using the SolexaQA package (Cox et al. 2010); reads were trimmed where Phred quality scores dropped below 17.

4.3.6. Use of publically available metagenomes from distant tundra and temperate grassland and agricultural habitats

Publically available metagenomes that were used for comparison purposes represent the 10-20 cm and 50-60 cm depths from Nome Creek, AK (NC; 200 km from CiPEHR site; Taş et al. 2014), active layer soil (30-35 cm depth) from Bonanza Creek, AK (BC; 100 km from CiPEHR site; Hultman et al. 2015), Toolike Lake LTER study site (TL; 530 km from CiPEHR site; Fierer et al. 2012), temperate steppe ecosystem in inner Mongolia, China (ZXM; Zhang et al. 2015), and agricultural soil from Urbana, IL (UIL; Orellana et al. 2014). All datasets were processed and analyzed as described above for CiPHER and KFFL datasets for consistency purposes, when possible. The Toolik Lake metagenome was omitted for comparisons of taxonomic composition and community complexity because the dataset size was comparatively smaller (<500 mbp). A summary of each site, including sampling depth, year of sampling, etc. is provided in Table C-2.

4.3.7. 16S rRNA analysis from amplicon PCR and shotgun-metagenome reads

Amplicon PCR 16S rRNA sequences longer than 190 bp (75% of expected length, 253 bp) after trimming were used for further analysis. QIIME was employed for the majority of downstream analysis (Caporaso et al. 2010b). 16S rRNA gene (16S) amplicon sequences were assigned to the sample they came from using a unique 12 bp barcode identifier, allowing for up to one mismatch. The Python script pick_de_novo_otus.py was used to cluster 16S sequences at 97% similarity (97% OTUs) with UCLUST (Edgar 2010). Representative sequences of each OTU were taxonomically identified with the RDP Classifier (Wang et al. 2007). Representative sequences were also aligned using PyNAST (Caporaso et al. 2010a), and a phylogenetic tree was constructed from this alignment using FastTree with default settings (Price et al. 2010**)**. Information on dataset quality and number of sequences used per sample is available in Table C-3.

The relative abundances of various prokaryotic taxa were also determined based on 16S sequences recovered in the metagenomes. Metagenomic reads were trimmed and sister reads were merged using the same approach as above. 16S sequences were identified by searching all merged and unmerged reads longer than 70 bp after trimming against the May 2013 release of Greengenes 16S ribosomal database (DeSantis et al 2006) pre-clustered at 97% identity, using blastn (BLAST+ version 2.2.29, options: -word_size 16 -outfmt 6 -task blastn -dust no -max_target_seqs 1) (Camacho et al. 2009). Only matches along the V4 region of the 16S sequence with bit-score ≥60, evalue <1E-7, and match length ≥70 bp were retained for analysis. 16S sequences then underwent open-reference OTU-picking against Greengenes database pre-clustered at 97%

(options: -m uclust_ref –s 0.97) (Edgar 2010) (similar to methods performed in Luo et al. 2014a). Representative sequences of each OTU were taxonomically identified with the RDP Classifier (Wang et al. 2007). Taxonomy abundances for each sample were determined by the number of sequences that assigned to OTUs corresponding to that taxonomic group, divided by the total number of reads that passed open-reference OTU-picking step (*i.e.,* that were used in clustering and not discarded). Euclidian distances of OTU tables were calculated, and the statistical significance of tundra vs. temperate topsoil OTU differences was determined by ANOSIM analysis. Results generalized at the phylum level were highly consistent between 16S rRNA gene analyses derived from 16S PCR amplicon sequences and 16S sequences recovered from metagenomic datasets.

4.3.8. Analysis of shotgun-metagenome short reads

Protein prediction of metagenomic sequences ≥100 bp after trimming was performed with FragGeneScan (Rho et al. 2010) (Illumina 1% error model). Resulting amino acid sequences were searched against Swiss-Prot database (The UniProt Consortium), using blastp (BLAST+ version 2.2.28) (options: –word_size 3 outfmt 6, cutoff: bit score >75, alignment length ≥25 amino acids, amino acid identity ≥40%) (Camacho et al. 2009). Corresponding Gene Ontology (GO) Annotations of functions and processes for each Swiss-Prot entry was obtained from http://www.uniprot.org/downloads (downloaded on December 04, 2014). Dataset quality and number of sequences used per sample is displayed in Table C-4.

A raw (not-normalized) counts table of genes and GO pathways (with sample metagenomes as columns and gene annotations or metabolic process categories as rows) was processed with the DESeq2 package (Love et al. 2014)

to identify differentially-abundant pathways between the two study sites and to generate log2 transformations of gene/process abundance ratios. The raw count data underwent a variance-stabilizing transformation, which is used for logarithmically distributed count data with low mean values that tend to have high variance. This transformation results in new values that have a relatively constant variance along the range of mean values and confers a reduced false positive rate for less abundant genes (Anders and Huber 2010). P-values were transformed to account for false discovery rate from multiple testing using Benjamini-Hochberg correction (adjusted p-values; Benjamini and Hochberg 1995). DESeq2 was also used to generate a sample-to-sample distance heatmap based on Euclidean distances derived from variance-stabilizing transformations of the raw count information, as well as to generate PCA plots based on the same transformations in order to visualize the overall effect of experimental covariates and batch effects.

4.3.9. Assembly and characterization of population bins

Initial assembly of metagenomic sequences was performed on each of the 11 CiPEHR Alaska datasets individually with CLC assembly package (http://www.clcbio.com) (options: -w 53). Resulting contigs ≥2 kbp were pooled together into one file and short metagenomic sequences were recruited to each contig to calculate the median coverage of each contig in each metagenome dataset (megablast with default; cut-off used: ≥90% of length of the query sequence, ≥98% nucleotide identity). Contigs ≥2 kbp were then binned using CONCOCT (Alneberg et al. 2014). Contigs that were not binned with CONCOCT were added to existing bins generated with CONCOCT if the unbinned contig matched a binned contig ≥99.7% nucleotide identity and ≥2 kbp alignment length

(using megablast with default settings). For each bin, re-assembly was performed using only short reads matching the binned contigs; matches ≥98% nucleotide identity and ≥100 bp length were assembled *de-novo* with Velvet (Zerbino and Birney 2008), including both merged and nonmerged paired-end reads, using odd-numbered word sizes between 85 and 99. These larger word sizes were chosen to increase assembly quality. Only contigs >1 kbp generated by this final assembly were included in the final bin; all contigs <1 kbp and reads and contigs from previous steps were discarded. CheckM was used to estimate completeness and contamination of each bin based on the recovery of 104 single copy universal bacterial genes (Parks et al. 2015).

Protein-coding genes of the assembled bins were predicted with MetaGeneMark (Zhu et al. 2010) and were searched against the Swiss-Prot database as described above (The UniProt Consortium). The sequence abundance of each bin was determined by the number of merged reads ≥200 bp from each metagenome that matched a population bin genome at ≥200 bp match length and ≥98% nucleotide identity. Closest ancestry was determined for certain organisms containing full or partial 16S and 23S rRNA sequences by searching contigs against SILVA small subunit (SSU) and large subunit (LSU) databases (v119 releases) (Quast et al. 2013) using megablast. A tree of phylogenetic relatedness based on 23S sequences was made for *Acidobacteria* bins 03, 06, 07, and 12 using *Ca. OP8* as an out-group, including several *Acidobacteria* isolates and uncultured organisms. Fasta sequences were aligned using MUSCLE algorithm (Edgar 2004) and PyNAST for tree building (Caporaso et al. 2010a). The resulting phylogenetic tree was visualized with FigTree V1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

4.3.10. Assessing intra-population diversity and biogeography

To assess whether the assembled bins were distributed over large geographical areas, trimmed reads from several publicly available metagenomes were searched against each population bin (reference) sequence using megablast. Matches ≥60% nucleotide identity and ≥70 bp length were retained. Fragment recruitment of these matches was performed to assess evenness and percent representation of each population bin in the corresponding metagenomes. Population structure was assessed using recruitment plots, essentially as described previously for marine metagenomes (Konstantinidis and DeLong 2008) and below.

To further validate the results of the recruitment plots, the full-length (3,211-4,462-bp) β subunit of the bacterial RNA polymerase (*rpoB*) gene for the population/bin of interest was used to recruit highly matching sequences from the metagenomes of the same or different sites. *rpoB*-encoding reads were identified using megablast. All matching reads were then used in a Velvet assembly to reconstruct partial or full-length sequences, when possible (*e.g.,* high enough coverage). Assembled *rpoB* sequences were aligned and then truncated to retain only the overlapping regions of all assembled *rpoB* sequences (overlapping region of CiPEHR original assembly and Nome Creek partial or full *de-novo* assembly). Using megablast, unassembled metagenomic reads matching the *rpoB* sequences in the alignment were identified from CiPEHR, Nome Creek, and Toolik Lake metagenomes, using as cut-off for a match ≥85 bp alignment and ≥98% nucleotide identity. Representative reads from each location were added to each existing alignment with MAFFT Multiple Sequence Alignment (options: mafft --anysymbol --addfragments –multipair) (Katoh and Frith 2012), and a tree was generated using Randomized Axelerated Maximum Likelihood, or RAxML,

version 730 (Stamatakis 2006). The resulting tree was visually inspected for the existence of a star-like phylogeny, indicating no population structure.

We developed a resampling approach to identify genes absent from *in-situ* populations in distant tundra habitats (NC, BC, or TL) that otherwise closely matched (>95% Average Nucleotide Identity [or ANI; Konstantinidis and Tiedje 2005], >90% genome representation) one of the 27 reference CiPEHR population bin assemblies. Population bin genomes were broken up into 500 bp segments *in-silico* and the average coverage for each segment was calculated independently, using recruitment of all available unassembled short sequences from the target metagenome. A skewed normal distribution was fit to these coverage values using the 'enveomics.R' package available for download at http://enveomics.blogspot.com. The parameters of the resulting skewed distribution were used to calculate the probability that a gene with zero or near-zero coverage (*i.e.,* no or few reads matching to that gene) could occur by chance (null hypothesis being that the gene was present in the population). 500 bp segments were treated as independent tests, and thus, p-values for each gene were adjusted by their corresponding lengths as follows: p-value ^ ([gene length] / 500 bp). Therefore, comparatively longer genes were more robustly assessed for presence/absence. This analysis was performed on all genes for selected comparisons, and p-values were adjusted to account for false discovery rate from multiple testing using Benjamini-Hochberg correction (Benjamini and Hochberg 1995)

4.3.11. Data availability

Assembled population bins, as well as raw shotgun metagenome and 16S rRNA gene sequences from CiPEHR, AK and KFFL, OK, are deposited in the

European Nucleotide Archive (http://www.ebi.ac.uk/ena/) under study no. PRJEB10725.

4.4. Results

4.4.1. Relative microbial community complexity of tundra soils

Using Nonpareil, a statistical tool that employs read redundancy to estimate the coverage of the microbial community achieved by a metagenomic dataset (Rodriguez-R and Konstantinidis 2014), a much more diverse community was observed in temperate soils compared to those from Alaskan tundra. The estimated sequencing depth required to sample 95% of the total extracted community DNA for each CiPEHR, AK soil sample was found to be 53.6±5.45 Gbp (mean±one standard deviation), 56.0±25.7 Gbp for Nome Creek 10-20 cm and 50-60 cm depth metagenomes, and 34.6±14.7 Gbp for Bonanza Creek active layer (~30-35 cm depth) metagenomes (Figure 4-1 and Table C-5). Using temperate soil metagenomes for contrast, an estimated 450±15.2 Gbp of sequencing depth would be required to sample 95% of sequencing richness for KFFL, OK soils, 281 Gbp for Urbana, IL (UIL) agricultural soil, and 215 Gbp for temperate steppe soil from inner Mongolia, China (ZXM). Based on a total of 139 Gbp and 62.2 Gbp available from all replicated metagenomes from CiPEHR and KFFL (merged sequences only), an estimated 91.8% of the combined 'site community' for CiPEHR and 62.2% for KFFL was sampled. Notably, the 'site community' represented by all 11 combined CiPEHR metagenomes *still* possessed less than half the complexity (or estimated sequence richness) estimated from a single KFFL metagenome. A higher level of diversity in the OK soil microbial community is further reflected in the 97% OTU rarefaction curve

from 16S PCR amplicon sequences, where the number of OTUs detected at OK is over two-fold greater than the number of OTUs observed at AK (Figure C-1).

4.4.2. Microbial community compositional differences between study sites

Using 16S PCR amplicon sequences prepared in parallel from KFFL, OK and CiPEHR, AK soil samples, most OTUs were shared between soil samples from the same site, but comparatively few OTUs were shared between the two sites. In particular, <6% of OTUs were shared (n = 766, from a total of 8,290 non-singleton OTUs present at OK and 6,292 at AK) and these differences were statistically significant (p-value < 0.001; ANOSIM) (Figure C-2). Further, by using metagenome-extracted 16S sequences to compare CiPEHR and KFFL datasets with those publically available, it was found that, on average, 72.5% of the OTUs from active layer Bonanza Creek and 74% of the OTUs from Nome Creek soils were detectable in other tundra locations (Figure 4-2). These results revealed that a high degree of OTUs are shared within the Alaskan tundra ecosystem vs. between ecosystem types (p-value < 0.01; ANOSIM), regardless of slight methodological differences between studies (DNA extraction procedure, handling, etc.). This trend is not observed when OTU affiliations are summarized into broad taxonomic groups, such as phylum-level evaluation (p-value > 0.1; ANOSIM).

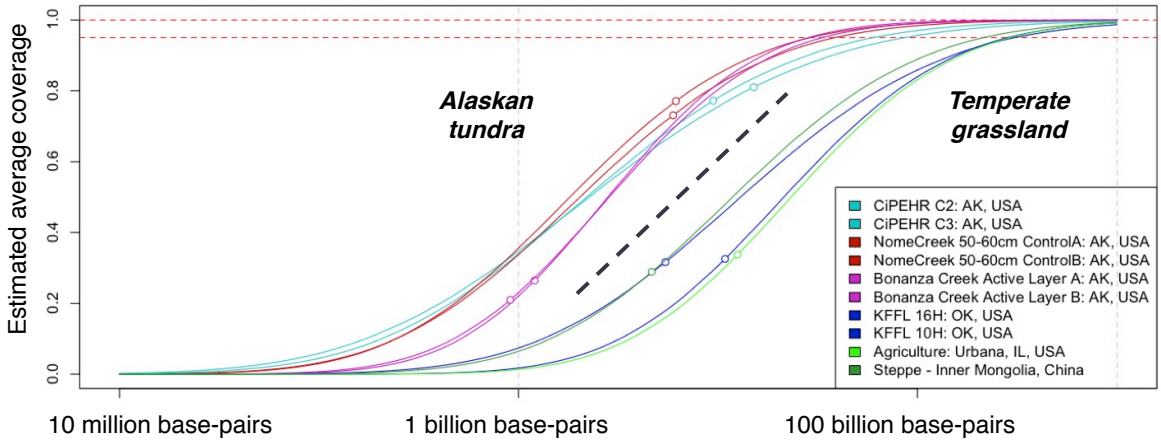


Figure 4-1: Curves representing soil microbial community complexity estimations as determined by Nonpareil. Nonpareil is a statistical tool that uses read redundancy to estimate dataset complexity and the amount of sequencing effort needed to achieve a desired level of coverage. Circles on curves represent the coverage of the actual sequencing depth for each dataset in relation to the entire curve (projection for complete coverage after the circle). Curves positioned on the right represent more sequence diverse metagenomes than curves positioned on the left.

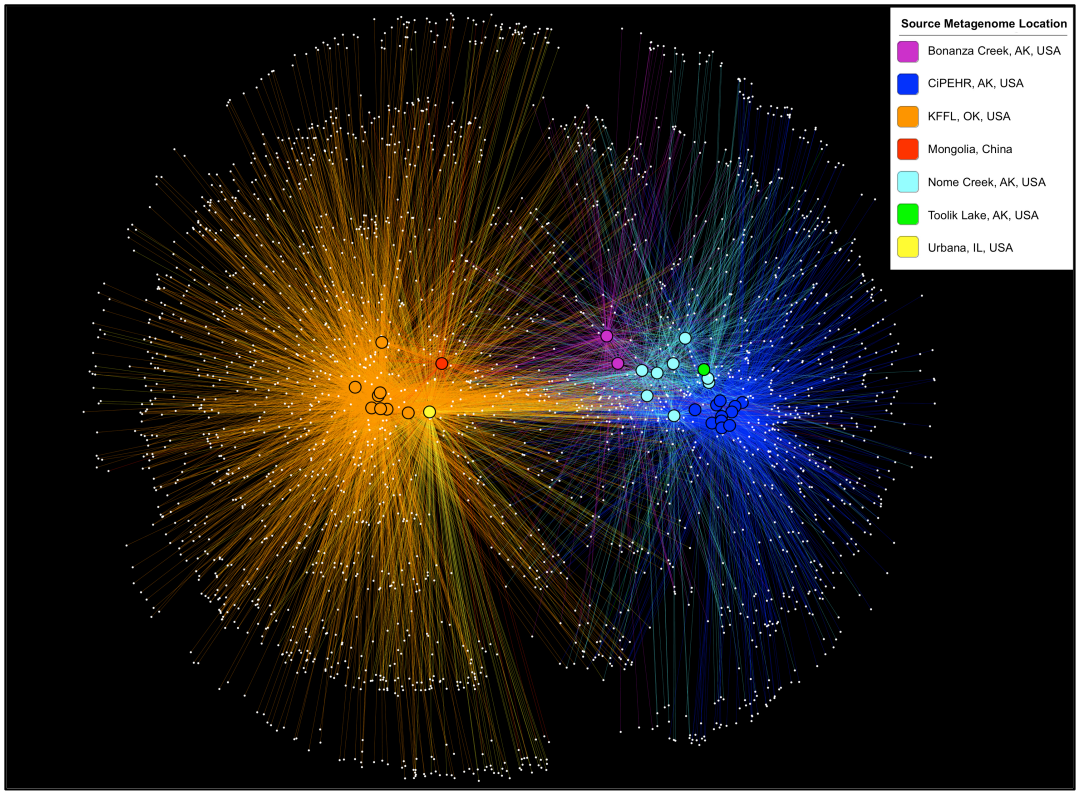


Figure 4-2: OTU sharing network based on 16S rRNA gene sequences from metagenomes. Samples are clustered (positioned) according to the presence and abundance of their shared OTUs (using make_otu_network.py, a QIIME script; Caporaso et al. 2010b). White dots represent OTUs, and a line connecting these dots to a bolded sample dot indicates that the OTU is present in that sample. The colors of lines and bolded sample dots correspond to the source location of each metagenome as provided in the key.

16S fragments recovered from metagenomes revealed several broad taxonomic groups that were common to either tundra or temperate ecosystems (*i.e.*, ubiquitous in one, absent from the other). Namely, methanogenic archaeal classes *Methanomicrobia and Methanobacteria* were present in all tundra locations (CiPEHR, BC, and NC) (0.22% of total community, on average), but were non-detectable in any temperate soil metagenome (KFFL, ZXM, and UIL). Other groups specific to tundra soils include *Chlorobi* class *Ignavibacteria*, *Bacteroidetes* class *Bacteroidia*, *Crenarchaeota* class *MCG*, *Nitrospirae* family *Thermodesulfovibrio* (obligate anaerobic sulfate reducer; Henry et al. 1994; Sekiguchi et al. 2008), *Chloroflexi* class *Dehalococcoides*, and phylum *Lentisparae*. Conversely, *Nitrososphaeraceae,* an archaeal family of ammonia oxidizers, was detected in all temperate soils (~1% abundance, on average), but was non-detectable in all tundra soils. *Nitrosomonas*, an ammonia oxidizing bacterium (Campbell et al. 2011), displayed patchy representation in OK 16S rRNA amplicon datasets, but was non-detectable in all tundra metagenomes or when using CiPEHR 16S rRNA amplicon sequences (Table C-6).

PCR amplicon sequences of the 16S rRNA gene (which allowed for a greater number of sequences per sample for analysis, relative to using 16S sequences derived from metagenomes) revealed that phylum *Acidobacteria* dominated the CiPEHR soil communities, representing 51.2%±11.62% of all members (mean±one standard deviation; Table C-7), followed by *Proteobacteria* (15.5%±6.2%), *Verrucomicrobia* (15.1%±4.4%), *Actinobacteria* (8.7%±2.8%), and *Bacteroidetes* (2.7%±4.8%). Oklahoma soils were instead dominated by *Proteobacteria* (29.2%±4.4%), *Verrucomicrobia* (20.9%±6.6%), *Acidobacteria* (20.2%±1.9%), *Actinobacteria* (9.2%±2.8%), *Bacteroidetes* (4.0%±1.0%), *Planctomycetes* (3.3%±1.2%), and *Firmicutes* (2.5%±1.7%). Less-dominant

phyla that were markedly more abundant at OK *versus* AK were *Cyanobacteria* (0.72% *vs*. 0.10%; OK mean *vs*. AK mean), *Chloroflexi* (1.80% *vs*. 0.41%), *Nitrospirae* (1.21% *vs*. 0.24%), *Crenarchaeota* (0.37% *vs*. 0.02%), *OP3* (0.13% *vs*. 0%), and TM6 (0.15% *vs*. 0.01%). AD3 (1.01% *vs*. 0.29%; AK mean *vs*. OK mean), and *Chlorobi* (0.48% *vs*. 0.04%) represented low abundant phyla that had a markedly greater mean abundance at CiPEHR compared to KFFL.

4.4.3. Metabolic comparison of tundra and temperate soils

Initial benchmarking with different read lengths indicated that a comparison including publically available metagenomes would not be robust given that sequences of different lengths presented inconsistent annotations; a finding consistent with those reported by Luo et al. 2014b. The publically available DNA sequences were all shorter than the 2x150 bp sequences available from CiPEHR, AK and KFFL, OK soil metagenomes. Thus, functional gene comparisons were limited to the CiPEHR, AK and KFFL, OK datasets.

Gene content dissimilarity derived from Euclidean distances of variance-stabilized data in DESeq2 (Figure C-3), displays high contrast between the CiPEHR, AK and KFFL, OK sites relative to dissimilarity between samples within a site. Analysis of functional genes from metagenomes involving the degradation of SOM revealed site-specific patterns for several processes. SOM degradation genes that were significantly more abundant in AK soil metagenomes compared to OK included those involved in the catabolism of chitin (79.3% percent higher abundance in AK relative to OK), cellulose (73.2%), simple carbohydrates (81.2%, on average, for monosaccharides; 40.1%, on average, for sugar acids; 68.6%, on average, for sugar alcohols), and lipids such as triglycerides and phospholipids (80.6% and 62.9%, respectively) (adjusted p-value <1e-3;

Negative binomial test with DESeq2; Figure C-4). These findings were also congruent with previous studies of Artic tundra soils that found high numbers of chitinase, sugar alcohol, and mono- and disaccharide degradation genes in metagenomes (Yergeau et al. 2010) and metatranscriptomes (Tveit et al. 2014). Genes related to the catabolism of lignin and phenolic compounds were 71.0% and 58.9% more abundant in OK metagenomes, respectively (adjusted p-value <1e-3; DESeq2; Figure C-5). Genes corresponding to the Wood–Ljungdahl carbon fixation pathway (carbon monoxide dehydrogenase/acetyl-CoA synthase subunit alpha and 5-methyltetrahydrofolate:corrinoid/iron-sulfur protein co-methyltransferase proteins), which are primarily used by acetate-producing bacteria and methanogens (Roberts et al. 1994, Matschiavelli et al. 2012), were 50-200 times more abundant in AK metagenomes, relative to OK (Table C-8).

While genes related to denitrification (N-loss from ecosystem) were more abundant at OK (*e.g.*, >3 times greater abundance of *nosZ* at OK), genes related to N-fixation (N-gain) were significantly more abundant in AK relative to OK datasets (46%, on average) (adjusted p-value <1e-3; DESeq2; Figure C-5). More specifically, NifH nitrogenase iron protein was more than 80 times more abundant in AK relative to OK metagenomes (Table C-8). SoxAX cytochrome complex subunit A, a protein responsible for sulfur oxidation, was much higher in abundance in OK metagenomes, whereas genes involved in sulfate reduction (primarily sulfate adenylyltransferase subunit proteins), hydrogen sulfide biogenesis and fermentation genes were more abundant in AK soils.

4.4.4. Contig binning and population reconstruction from Alaska soil metagenomes

Contig binning resulted in the recovery of twenty-seven population bins, twelve of which were >80% complete based on the recovery of 104 single-copy universal bacterial genes using CheckM (Parks et al. 2015), with 0%-1.7% contamination (Table 4-1 and Table C-9). Fourteen of the bins had an N50 (*i.e.*, the length that >50% of the assembly is in contigs of this length or longer) between 20,203 bp and 218,163 bp. Collectively, the 27 assembled population bins recruited 7.5% of all reads, on average (9.2% max), for the 11 metagenomes (≥98% nucleotide identity; ≥200 bp alignment).

Table 4-1: Summary statistics for the most complete population bins. Assembly statistics, including N50 (*i.e.*, the length that >50% of the assembly is in contigs of this length or longer), the length of the longest contig, the total length of all contigs combined, and the number of contigs comprising a bin, as shown. All contigs were >1 kbp long. CheckM was used to assess each bin for completeness, evaluating the presence of 104 single copy bacterial genes. The number in each column represents the number of instances each of the 104 bacterial single copy genes was found in each bin (*i.e.,* 98 single copy genes were found once in Bin 01). Bins highlighted in green indicate those detected in distant tundra locations (100-530 km from CiPEHR, AK) at >2X representation of the genome. Single copy marker genes that occur more than once served as an indicator of contamination – *i.e.,* that the assembly is combined with sequences from a more than one organism.

	POPULATION BIN CONTIGS				SINGLE COPY GENES			CHECKM RESULTS	
Bin ID	n50 (bp)	Longest (bp)	Total Length (bp)	Number Contigs	0	1	2	Complete-ness	Contamin-ation
1	87585	258416	4341484	85	5	98	1	94.0%	1.7%
2	44196	144885	4793253	228	21	83	0	89.7%	0.0%
3	167654	415574	5822639	110	4	100	0	95.7%	0.0%
4	36041	146424	2468837	156	14	90	0	81.0%	0.0%
5	20203	76412	2053920	189	12	92	0	83.5%	0.0%
6	54187	226195	3684989	154	2	102	0	97.4%	0.0%
7	142525	388009	5910590	149	12	92	0	89.8%	0.0%
8	65109	167976	3268372	73	3	100	1	95.7%	1.7%
10	12295	50046	2301754	289	18	85	1	81.3%	1.7%
11	51987	197629	5172412	191	2	102	0	98.8%	0.0%
20	13258	53656	3231835	375	31	71	2	83.1%	1.0%
22	27045	163169	4115219	235	30	74	0	74.4%	0.0%
27	26244	147429	8218361	566	3	101	0	94.8%	0.0%

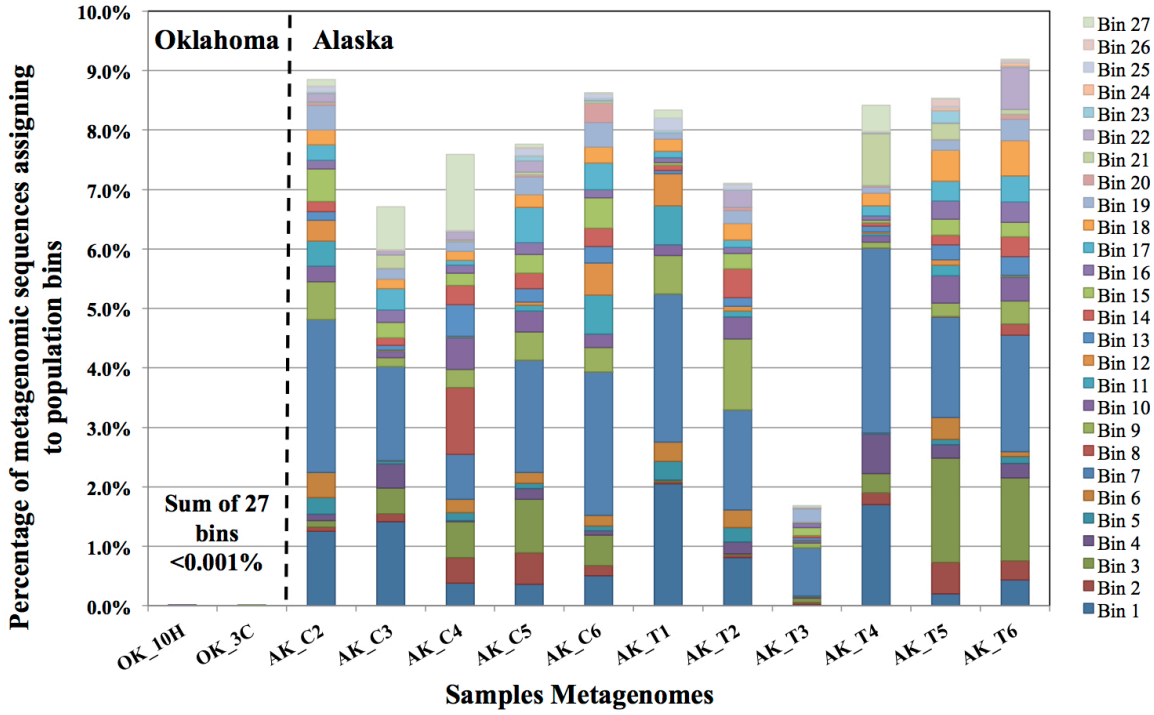


Figure 4-3: Abundance of individual populations (bins) in various Alaskan tundra metagenomes determined by read mapping. Results were obtained by the number of sequences that matched a population genome at ≥200 bp and ≥98% nucleotide identity for CiPEHR metagenomes (≥90 bp and ≥98% nucleotide identity for metagenomes at other locations), divided by the number of reads used as query (all merge-able paired-end sequences ≥200 bp at CiPEHR; sequences ≥90 bp at other sites). Bin abundance is given for two KFFL, Oklahoma metagenomes for contrast. Taxonomic affiliation is provided for bins with full or partial 16S ribosomal sequence, along with the average amino acid identity (AAI) to the corresponding genome. *Nucletide identity given corresponds to the 23S sequence because the 16S gene was not available. **AAI is derived from comparison to a genome that is not the same as the 16S match because a genome sequence was not available for these references; instead, the next closest match was used.

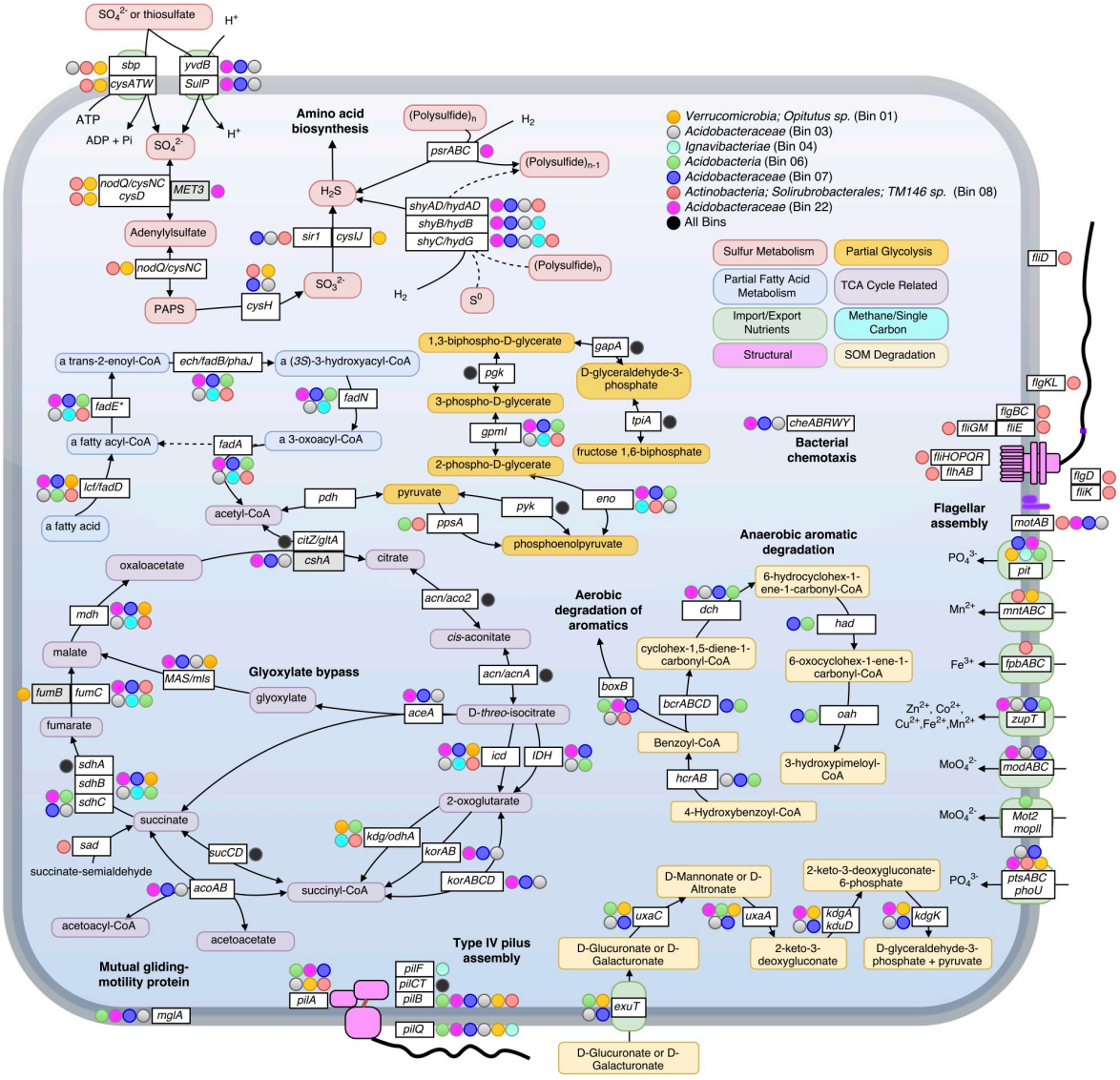


Figure 4-4: Metabolic pathways identified in seven assembled population bins. Selected metabolic transformations and other cellular activities are represented by both the name of the gene known to encode the protein enzyme for them as well as the substrate of the enzyme. Protein enzymes were identified by searching predicted proteins against Swiss-Prot database, as well as by searching against close representatives of selected organisms: *Acidobacteria Ca. Solibacter usitatus Ellin6076* and *Ca. Koribacter versatilis Ellin345* for bins

03, 06, 07, and 22; *Chlorobi Ignavibacterium album JCM 16511* for bin 04; *Verrucomicrobium Opitutus terrae PB90-1* for Bin 01. In all instances besides *zupT*, *mglA*, *phoU*, *sad*, and *pit,* pathways were only displayed if all or most genes of an operon involved in the same pathway/process were detected as present. Colored circles alongside genes indicate that the bin assigned to that color (see legend on top right) encoded the gene (blastp cutoffs: e-value <1e-20, bitscore >100). A black-filled circle indicates that all seven bins in the upper right box possessed the accompanying gene.

We have chosen to highlight seven of the most complete population bins below (see also Table 4-1 and Figures 4-3 and 4-4). Bin 01 contained full-length 16S and 23S rRNA gene sequences matching at 96% and 93.1% nucleotide identity, respectively, to an *Opitutus sp.* (phylum *Verrucomicrobia*) isolate from an anoxic region of rice paddy soil (Chin et al. 2001) (Table C-10). Protein annotation of this bin revealed the presence of genes related to methanotrophy, including methylamine and methanol pathways (*mtaB, mtbA, mttC1*). Bin 01 also contained genes involved in assimilatory sulfate reduction and transport (*sbp*, *cysATW, cysDGHIJK, nodQ*), the degradation of a variety of organic compounds, including L-arabinose, xylan/xyloglucan, D-glucuronate, ribose, cellulose, and rhamnose, and the metabolism of molecular H_2 (*hndACD, hoxHY, hypABCDEF, hupE*).

Bins 03, 07, 11, 12, and 22 were closely related (>80% ANI) and together formed a novel monophyletic clade within the family *Acidobacteraceae* based on partial 23S rRNA gene sequences (Figure C-6). Bin 03 contained 100/104 single-copy bacterial genes, and partial 16S and 23S rRNA gene sequences (338 bp

and 568 bp, respectively; Table C-10). The 23S sequence matched family *Acidobacteriaceae* (94.4% nucleotide identity). Bin 07, possessed a partial (358 bp) 23S sequence that matched *Granulicella mallensis MP5ACTX8* (also *Acidobacteriaceae*) at 94.7% nucleotide identity, which was previously isolated from tundra soil in Finland (Männistö et al. 2012). Bin 03 and 07 shared many functional genes; including Citric Acid Cycle (TCA) with glyoxylate bypass (*icl*, *MAS*), in addition to genes required for 'typical prokaryotic TCA metabolism'. Both bins possessed genes for trimethylamine utilization and degradation (*mttBC*, *etfABD*, *mauG*) and an ammonium ion transporter on the same operon (*nrgA*), and several genes involved in the degradation of many SOM constituents, such as rhamnose import and catabolism (*rhaT, rhaB, rhaM*), degradation of fatty acids (*lcf*, *fadADEN*, *ech*), genes for catabolism of succinate (*sdhABC*), formate (*fdoGI*, *fdhE*, *fdnHG*), and several other genes for the catabolism of xylan, xyloglucan, cellulose, arabinan, and chitin and its derivative, N-acetylglucosamine. Bin 03 also contained genes for coenzyme B–coenzyme M heterodisulfide reductase (*hdrABC*) and both bins contained genes for other methane-related coenzyme activity (*cofDEGH*). Bin 07 possessed genes for anaerobic and aerobic degradation of aromatic compounds (*hcrAB*, *bcrABCD*, *dch*, *had*, *oah*, *boxB*). On average, 1.95% of metagenomic reads matched bin 07 at ≥98% nucleotide identity, making it the most abundant of the assembled populations at the AK site (Figure 4-3).

Bin 22 shared many metabolic strategies with bins 03 and 07. It also contained genes for TCA cycle with glyoxylate bypass (*icl*, *MAS*), as well as genes for the degradation of carbohydrates and fatty acids. All three bins possessed genes for bacterial chemotaxis (*cheABRWY*), motility proteins

(MotAB), and along with bin 06, possessed mutual gliding protein (MglA), but none possessed genes necessary for flagellar assembly.

Bin 06 possessed 102/104 single-copy genes, and contained a large contig ending in a partial 23S ribosomal sequence matching a representative of "*Candidate Division OP8*" and an uncultured *Acidobacteria* at 79.9% and 80.2% nucleotide identity, respectively (Table C-10). Therefore, it probably represents a group more distant to known organisms than other assembled bins (Figure C-6). This bin contained genes for rhamnose, cellulose, and D-glucarate catabolism, hydrogen metabolism (*hyfBDEF*), sulfoxide reductase (*yedYZ*), and a wide variety of genes for the degradation of aromatic compounds (*hcrAB*, *bcrABCD*, *dch*, *had*, *oah*, *boxBC*, *nicAB, catIJ, mphP*).

Bin 04 possessed a partial 23S rRNA gene sequence (760 bp) matching to *Ignavibacteria* (phylum *Chlorobi*), and also contained Coenzyme B–Coenzyme M heterodisulfide reductase genes (*hdrABC*), which can be involved in anaerobic methanotrophy, sulfur reduction, or fermentative pathways. Bin 04 possessed only the beta and gamma subunits of the sulfyhydrogenase I complex (*hydGB*) implying that it likely transforms elemental sulfur into hydrogen sulfide with molecular H_2, or performs the reverse reaction. Bin 08, which contained full 16S and 23 rRNA gene sequences, best matching actinobacterial family *Solirubrobacterales* (99.1% and 93.4% nucleotide identity, respectively), comprised of 73 contigs and contained 101/104 single copy genes. This genome possessed motility proteins (MotAB), as well as genes required for flagella assembly (*fliEGHKMOPQR*, *flhAB*, *flgBCL*). It also encoded genes for sulfate binding, import, and assimilation (*sbp*, *cysATW*, *nodQ*, *cysH*, *sir1*), polysulfide reduction using molecular H_2 (*shyACD*), trimethylamine dehydrogenase (*tmd*),

and genes related to the catabolism of fatty acids, chitin, oligopeptides, lipopolysaccharides, mannose, galactose, acetate, succinate and pyruvate.

None of the twenty-seven bins possessed denitrification genes *narG, nirS*, *norB*, or *nosZ*. Bin 20 was found to encode all genes necessary for nitrogen fixation (*nifABDEHJKNQSV1WXZ*). Bin 20 also contained a contig ending in a partial 16S ribosomal sequence matching an *Acidobacteria* representative at 100% nucleotide identity (169 bp match) and 19/21 ribosomal proteins best matched to *Acidobacteria Ca. Solibacter* or *Ca. Koribacter*.

4.4.5. Population distribution in other high latitude soils

Read recruitment plots against the assembled population bin genome sequences revealed that all bins represented sequence-discrete populations, with short reads matching population bin sequences mostly at 99%-100% nucleotide identity, evenly across the entire genome (examples presented in Figure C-7). These findings revealed that soil microbial communities are composed of discrete populations, at least for the abundant fraction that is robustly assessed by metagenomics. These findings echoed results reported previously for other habitats (Caro-Quintero and Konstantinidis 2012). Further, all 27 population bins were detected in more than one CiPEHR metagenome (Figure 4-3). Surprisingly, several bins matched reads in other publicly available tundra soil metagenomes (>98% nucleotide identity and even coverage across the genome sequence), indicating that these populations are widespread in high latitudes and revealing that long-lived, sequence-discrete, tractable populations are present in at least some soils. Specifically, datasets from the Nome Creek area (Taş et al. 2014), located ~200 km Northeast from the CiPEHR site, contained five of the assembled bins at high abundance (*i.e.*, 2-10X coverage of

population bin with unassembled short sequences) (Figure 4-3, 4-5 and C-8). Metagenomes containing these populations originated from surface (10-20 cm) and middle depth (50-60 cm) samples, which undergo seasonal thawing, similar to the CiPHER site. Metagenomes representative of the permafrost layer (90-100 cm) did not contain any of the assembled bins from CiPEHR. Bin 07, the most dominant population bin in CiPEHR and Nome Creek site metagenomes, was also found in active-layer soil from Bonanza Creek, AK (~100 km Northeast from CiPEHR) and in a soil metagenome from Toolik Lake, AK (530 km North of the CiPEHR site; Figure 4-5), along with bins 09 and 10 (Figure C-8). Because fragment recruitment using Marcell Experimental Forest soil metagenomes (SPRUCE experiment in Northern Minnesota; Lin et al. 2014) revealed the presence of a distinct acidobacterial population closely related to bins 03, 07, and 11 (~82% ANI), *de-novo* assembly was performed on these metagenomes in order to recover the corresponding population bin(s) (Figure C-9).

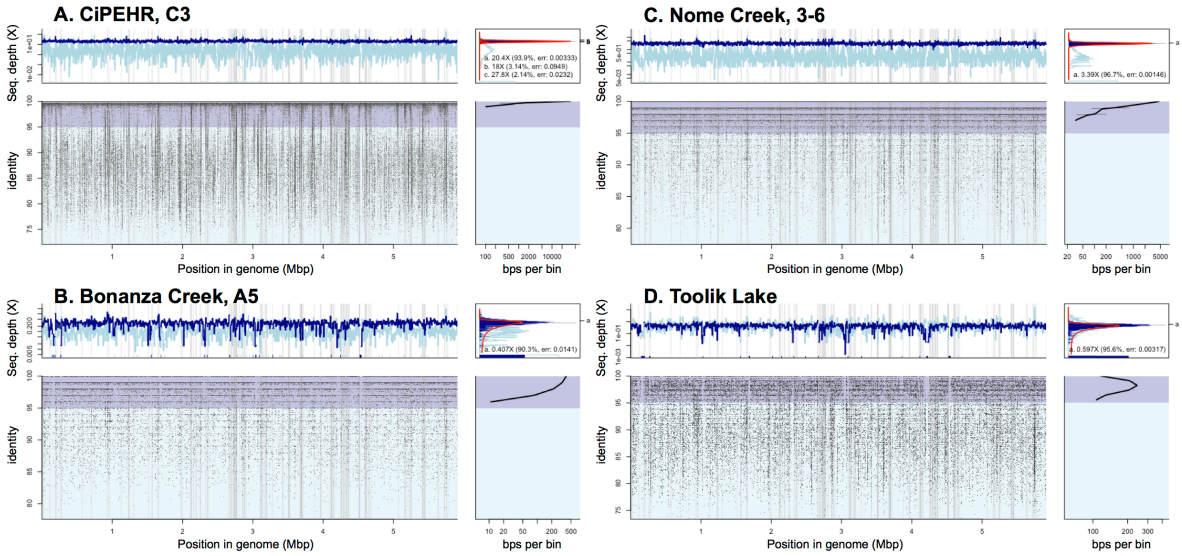


Figure 4-5: Read recruitment plot of population bin 07 at four distinct tundra locations. Diagrams expressing coverage of population bin 07 (assembled from CiPEHR, AK metagenomes) in metagenomic datasets from four distinct geographic locations, each separated by 100-530 km distances. All short reads from select sample metagenomes were searched against all contigs of the bin in a megablast search. The position histogram (top left of each) displays the average coverage of each base position, determined by a 1,000 bp window. The dark blue histogram represents the coverage at each base position determined from reads matching ≥70 bp in length and ≥95% nucleotide identity, while the light blue represents the coverage determined from reads matching ≥70 bp in length and <95% nucleotide identity. An even coverage across the entire contig from matching reads that are high identity (>95%) is indicative of a high quality contig from a single population. The recruitment plot (bottom left of each) shows where individual metagenomic reads matched to the population bin and the identity (%) of the match. The ID histogram (bottom right) displays the total

number of short read-derived base-positions at given percent identities. Note that in all cases shown, a sequence-discrete population represented by reads showing high nucleotide identity to the reference population genome sequence (typically >95% nucleotide identity) and even coverage across the length of the reference sequence is obvious.

4.4.6. Genomic adaptation to fire disturbance at Nome Creek, Alaska

Bins 07 and 09 were represented by ~3X coverage or more in both fire-impacted (soil sampled seven years following a fire disturbance event) and control (undisturbed) soil metagenomes from Nome Creek, AK (Taş et al. 2014). The adequate coverage of the CiPHER bins obtained in these samples, allowed for the investigation, at a fine resolution level, of the genomic adaptations likely occurring as a result of the fire-related environmental changes. Briefly, fire disturbance at this site resulted in a large reduction in SOM and soil moisture content, complete thawing of permafrost, and several years of elevated temperatures relative to adjacent undisturbed soils. All genes belonging to population bin 07 were represented in metagenome NC3-6 (50-60 cm depth, undisturbed; hence, complete representation of the population bin), with an average nucleotide identity of 99.6% using all recruited matches resulting from megablast (cutoffs: matching length >70 bp, nucleotide identity >60%) (Figure 4-5C). Comparatively, the average nucleotide identity of matching sequences in NC2-6 (50-60 cm depth, fire-impacted) was 99.1%, and 130 genes of the reference genome (2.5% of the total) were absent from the *in-situ* population (adjusted p-values <0.025). Similarly, recruitment of short reads from metagenome NC3-6 revealed near-complete representation of population bin 09

genes with an estimated ANI of 99.5%, whereas 209 genes of the reference genome were absent from the NC2-6 *in-situ* population, and an ANI of 98.9% estimated from the recruited sequences (adjusted p-values <0.025). Absent genes included those involved in transcriptional regulation (9% and 13% in bins 07 and 09, respectively), carbohydrate transport and metabolism (18% and 12%), amino acid transport and metabolism (13% and 14%), and nutrient transport (8% and 11%), among other functions and hypothetical genes. Examples of selected contigs from bins 07 and 09 with high gene absence in fire-impacted soils are displayed in Figure C-10 (read recruitment from undisturbed soils are also shown for contrast).

4.5. Discussion

4.5.1. Assessment of microbial biogeography reveals conserved, pervasive bacterial populations in the Alaskan tundra ecosystem

The high degree of microbial community heterogeneity often observed in soil systems remains poorly understood. Here, we reported the detection of discrete bacterial populations, several of which -but not all- were also found in locations that are separated by distances of up to 530 km (Figures 4-3, 4-5 and C-8). These findings were consistent with previous studies that reported a large degree of shared genera between tundra locations separated by up to 70 km (Frank-Fahle et al. 2014). The patchy representation of certain population bins within and between study sites uncovers an inherent feature of this heterogeneity, *i.e.,* although soils within close proximity (cm to m distances) may not share certain microbial populations, these populations can be shared by soil samples separated by several hundred kilometer distances. These findings

reveal that micro-variable soil conditions and niche space might be shared across large geographic regions in the tundra ecosystem, but not uniformly. The *Acidobacteria* bin 07, which possessed metabolisms for elemental hydrogen, elemental sulfur, sulfate, methanotrophy, TCA cycle with glyoxylate bypass, and catabolism of fatty acids, carbohydrates, and recalcitrant SOM such as chitin and aromatic compounds, was the most dominant population within the CiPEHR site metagenomes (Figure 4-3), and was also found, at high relative abundance (Figure 4-5), in metagenomes from three other tundra locations, all separated by ≥100 km. These findings revealed a successful generalist population, capable of thriving in many distinctive locations over broad geographical distances, and also suggests that microbial species conservation may dominate over high habitat specificity for certain soil prokaryotes. Furthermore, bins 03, 07, 11, 12, and 22 comprise a monophyletic group of closely related populations with >97% ribosomal gene sequence similarity (Figure C-6) or high ANI (~80% or greater). A close relative to this group (82.2% ANI to bin 11; Figure C-9) was assembled directly from Marcell Experimental Forest soil metagenomes (~3,780 km from CiPEHR; Lin et al. 2014), and recruited up to 6.8% of all metagenomic reads from its source metagenome. This finding underscores the high dominance of this relatively narrow (phylogenetic) *Acidobacteria* clade across large geographic distances (*i.e.,* 3,780 km, or more).

Previous studies have demonstrated the rapid response of active layer tundra soil microbes to elevated temperatures in the laboratory (Mackelprang et al. 2011), and here, we demonstrated the tempo and mode of adaptation of tundra bacterial populations to environmental perturbations such as prolonged fire events. In this case, adaptation took place by altering 2-3% of the genome within a period of seven years or less. This is quite remarkable for soils microbes,

especially those in low-temperature high latitudes, which are traditionally viewed as slow growers (Price and Sowers, 2004). Many of the genes absent from *in-situ* populations of fire-impacted soils (but present intact in undisturbed soils) were assignable to pathways related to the transport and catabolism of simple SOM substrates and gene regulation, implying that such disturbance events may have long lasting effects on soil C cycling. Taş and colleagues (2014) demonstrated that for middle layer depths (50-60 cm; same depth assessed in our own comparison) subject to fire, genes for simple carbohydrate metabolism decreased in abundance; here, we demonstrate that at least part of this change was attributable to the genomic alteration of existing, dominant microbial populations. Furthermore, the loss of transcriptional regulatory genes may reflect an alleviation of prior metabolic constraints (*e.g.,* due to increased temperature or decreased moisture content in this case). Consistent with these findings, Coolsen and Orsi (2015) demonstrated that transcriptional regulation and signal transduction represented a large category of genes that were differentially regulated between thawed or frozen permafrost soils. Our cross-site comparison exemplified the usefulness of combining datasets from multiple studies and showed that sequences and genomes recovered from soil can be meaningfully combined with datasets representing the same ecosystem elsewhere, at least in this Alaskan tundra ecosystem.

4.5.2. Contrasting taxonomic composition and functional gene content in tundra and temperate topsoil

The assembled population bins were highly representative of both the taxonomic and functional composition of the whole-community in CiPEHR, AK soils, epitomizing several pathways that were more abundant at tundra relative to

temperate metagenomes. Even though the bins were comprised of organisms from several phyla, many metabolic pathways were nearly ubiquitous amongst the bins. Namely, several organisms contained genes indicative of methane or methylamine metabolism(s). The majority of populations encoded genes for the formation or catabolism of gaseous hydrogen. Many also contained pathways for the degradation of SOM, capable of degrading compounds of varying recalcitrance, *i.e.*, chitin, cellulose, cellobiose, xylan, mannan, xylose, β-galactosides, raffinose, arabinose, aromatics, etc., several of which are highly expressed or abundant in tundra soils, as found previously (Yergeau et al. 2010; Tveit et al. 2014). A large portion of the latter compounds is apparently plant-derived, revealing the tight coupling between aboveground plant and belowground microbial communities in the tundra ecosystem. While genes responsible for the degradation of more labile carbohydrates are in greater relative abundance in CiPEHR metagenomes, it is clear that genes related to the degradation of many recalcitrant compound categories, such as phenolics and lignin, are more abundant in KFFL, OK metagenomes compared to CiPEHR (Figure C-4 and C-5), consistent with the differences in aboveground plant communities between the two sites. Further, the presence of genes recognized as hydrogeno- and methano-trophic in annotated short reads and assembled population bins, along with the absence of archaeal population bins and low overall methanogen abundance determined by 16S rRNA community analysis, suggests that methane production likely occurs at deeper layers in CiPEHR sites and migrates upwards towards the surface, where it is either released to the atmosphere or consumed by bacteria. Consistent with these interpretations, a recent study using metagenomes which included several of the assembled bins from CiPHER according to our analysis, displayed an overall reduced archaeal

presence in active layer compared to permafrost layer soil depths (Taş et al. 2014).

Previous studies have shown that permafrost ecosystems are nitrogen-limited (Shaver and Chapin, 1980; Martineau et al. 2010; Wild et al. 2014) and this was reflected in both our own Illumina datasets as well as the 454 datasets determined by Yergeau and colleagues (2010) for permafrost soils. For instance, although evidence of nitrate/nitrite transport was present, none of the binned populations displayed a significant role in the denitrification process. It is possible that decreased denitrification potential (*i.e.*, processes ultimately leading to nitrogen loss from the ecosystem) is a community adaptation to limiting N-availability, given that the genetic potential for this activity is lower at CiPEHR relative to KFFL, even though other metabolic pathways at CiPEHR are indicative of anaerobic conditions. N-limitation likely plays a role in constricting soil organic matter degradation in permafrost ecosystems. If N-cycling in these systems was altered due to anthropogenic related changes to the ecosystem (*e.g.*, agriculture practices and climate change), it is necessary to understand how this will relate to greenhouse gas emissions (climate feedbacks), *i.e.,* an increase in N-availability would stimulate aboveground plant communities and act as a C-sink or would increase denitrification activity resulting in acceleration of N_2O (a potent greenhouse gas) emissions. Furthermore, previous research demonstrated that addition of organic N to active layer tundra soil high in SOM content resulted in a 2-3-fold increase in SOM decomposition (Wild et al. 2014). Notably, one *Acidobacteria* (bin 20) possessed a complete assortment of nitrogen fixation genes, which is to the best of our knowledge, the first report of nitrogen fixation for a member of phylum *Acidobacteria*.

4.5.3. Relative taxonomic and functional complexity of tundra and temperate soil communities

Metagenomes representing temperate topsoils displayed greater taxonomic and functional diversity than those from Alaskan tundra. The lower diversity in tundra may be related to longer generation times, restrictive metabolic conditions (*e.g.,* O_2 restriction or lower temperature), and/or lower diversity of plant-derived organic carbon. Recent studies have shown that a constraint on diversity for certain functional traits results in decreased ecosystem functioning (Salles et al. 2012, Singh et al. 2014). Hence, changes in the diversity of functional genes relevant to SOM degradation may constrain or promote the rate at which these soils systems respond to environmental perturbations. Accordingly, the tundra microbial communities are presumably more vulnerable (less robust) to environmental change compared to their temperature counterparts.

4.6. Summary and future work

The degree to which microbial variability differs between ecosystems is essential for research efforts endeavoring to holistically understand soil microbial ecology in the context of ecosystem functioning and how environmental changes affect microbial activities and services. Our analyses revealed several taxa, at different resolution levels, *i.e.*, OTUs, taxonomic clades, or individual populations, that are ubiquitous, and frequently among the most abundant members of the corresponding communities, in the active layer of Alaskan tundra soils. Thus, their relative contribution to various ecosystem functions is expected to be high. These assembled bins and DNA sequences provided here represent potential

model organisms for future in-situ experimental manipulations; for instance, through the design of population-specific PCR for assessing gene transcript (activity) level, allowing potential linking of methane, nitrogen, SOM, etc. turnover to individual populations.

4.7. Acknowledgements

This research was supported by the U.S. Department of Energy (award DE-SC0004601).

4.8. References

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ et al. (2014). Binning metagenomic contigs by coverage and composition. Nature Methods 11: 1144-1146. doi: 10.1038/nmeth.3103

Anders S, Huber W. (2010). Differential expression analysis for sequence count data. Genome Biology 11. doi: 10.1186/gb-2010-11-10-r10

Benjamini Y, Hochberg Y. (1995). Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289-300.

Campbell MA, Chain PSG, Dang H, El Sheikh AF, Norton JM, Ward NL et al. (2011). Nitrosococcus watsonii sp. nov., a new species of marine obligate ammonia-oxidizing bacteria that is not omnipresent in the world's oceans: calls to validate the names 'Nitrosococcus halophilus' and 'Nitrosomonas mobilis'. Fems Microbiology Ecology 76: 39-48. doi: 10.1111/j.1574-6941.2010.01027.x

Caro-Quintero A, Konstantinidis KT. (2012). Bacterial species may exist, metagenomics reveal. Environmental Microbiology 14: 347-355. doi: 10.1111/j.1462-2920.2011.02668.x

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al. (2009). BLAST plus : architecture and applications. Bmc Bioinformatics 10. doi: 10.1186/1471-2105-10-421

Caporaso, JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6(8): 1621-1624. doi: 10.1038/ismej.2012.8

Caporaso, JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, and Knight R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America 108 Suppl 1: 4516-4522. doi:10.1073/pnas.1000080107

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335-336. doi: 10.1038/nmeth.f.303

Chin KJ, Liesack W, Janssen PH. (2001). Opitutus terrae gen nov, sp nov, to accommodate novel strains of the division 'Verrucomicrobia' isolated from rice paddy soil. International Journal of Systematic and Evolutionary Microbiology 51: 1965-1968.

CLC Assembly Cell 4.2.2. http://www.clcbio.com

Coolen MJL, Orsi WD. (2015). The transcriptional response of microbial communities in thawing Alaskan permafrost soils. Frontiers in Microbiology. 6:197. doi:10.3389/fmicb.2015.00197.

Cox MP, Peterson DA, Biggs PJ. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. Bmc Bioinformatics 11. doi: 10.1186/1471-2105-11-485

Curtis TP, Sloan WT, Scannell JW. (2002). Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America 99: 10494-10499. doi: 10.1073/pnas.142680199

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72: 5069-5072. doi: 10.1128/aem.03006-05

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797. doi: 10.1093/nar/gkh340

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461. doi: 10.1093/bioinformatics/btq461

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences of the United States of America 109: 21390-21395. doi: 10.1073/pnas.1215210110

Frank-Fahle BA, Yergeau É, Greer CW, Lantuit H, Wagner D. (2014). Microbial functional potential and community composition in permafrost-affected soils of the NW Canadian Arctic. PLoS One. Jan 8;9(1):e84761. doi: 10.1371/journal.pone.0084761

Grosse G, Harden J, Turetsky M, McGuire AD, Camill P, Tarnocai C et al. (2011). Vulnerability of high-latitude soil organic carbon in North America to disturbance. Journal of Geophysical Research-Biogeosciences 116. doi: 10.1029/2010jg001507

Handelsman, J, J Tiedje, L Alvarez-Cohen, M Ashburner, I Cann, E Delong et al. (2007). The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. Washington, DC, The National Academies Press.

Heimann M, Reichstein M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451: 289-292. doi: 10.1038/nature06591

Henry EA, Devereux R, Maki JS, Gilmour CC, Woese CR, Mandelco L et al. (1994). Characterization of a new thermophilic sulfate-reducing bacterium -

Thermodesulfovibrio yellowstonii, gen-nov and sp-nov - its phylogenetic relationship to Thermodesulfobacterium commune and their origins deep within the bacterial domain. Archives of Microbiology 161: 62-69. doi: 10.1007/bf00248894

Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, Blazewicz SJ et al. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature 521: 208-+. doi: 10.1038/nature14238

Jorgenson MT, Romanovsky V, Harden J, Shur Y, O'Donnell J, Schuur EAG et al. (2010). Resilience and vulnerability of permafrost to climate change. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 40: 1219-1236. doi: 10.1139/x10-060

Katoh K, Frith MC. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. Bioinformatics 28: 3144-3146. doi: 10.1093/bioinformatics/bts578

Konstantinidis KT, Tiedje JM. (2005). Towards a Genome-Based Taxonomy for Prokaryotes. Journal of Bacteriology. 187(18):6258-6264. doi:10.1128/JB.187.18.6258-6264.2005.

Konstantinidis KT, DeLong EF. (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. Isme Journal 2: 1052-1065. doi: 10.1038/ismej.2008.62

Lawrence DM, Slater AG, Swenson SC. (2012). Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4. Journal of Climate 25: 2207-2225. doi: 10.1175/jcli-d-11-00334.1

Lee H, Schuur EAG, Inglett KS, Lavoie M, Chanton JP. (2012). The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. Global Change Biology 18: 515-527. doi: 10.1111/j.1365-2486.2011.02519.x

Lin X, Tfaily MM, Green SJ, Steinweg JM, Chanton P, Imvittaya A et al. (2014). Microbial Metabolic Potential for Carbon Degradation and Nutrient (Nitrogen

and Phosphorus) Acquisition in an Ombrotrophic Peatland. Applied and Environmental Microbiology 80: 3531-3540. doi: 10.1128/aem.00206-14

Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15. doi: 10.1186/s13059-014-0550-8

Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K et al. (2014). Soil Microbial Community Responses to a Decade of Warming as Revealed by Comparative Metagenomics. Applied and Environmental Microbiology 80: 1777-1786. doi: 10.1128/aem.03712-13

Luo C, Rodriguez-R LM, and Konstantinidis KT. (2014). MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic. Acids Research. 42 (8): e73 first published online March 3, 2014 doi:10.1093/nar/gku169

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ et al. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480: 368-U120. doi: 10.1038/nature10576

Männistö MK, Rawat S, Starovoytov V, Haggblom MM. (2012). Granulicella arctica sp nov., Granulicella mallensis sp nov., Granulicella tundricola sp nov and Granulicella sapmiensis sp nov., novel acidobacteria from tundra soil. International Journal of Systematic and Evolutionary Microbiology 62: 2097-2106. doi: 10.1099/ijs.0.031864-0

Martineau C, Whyte LG, Greer CW. (2010). Stable Isotope Probing Analysis of the Diversity and Activity of Methanotrophic Bacteria in Soils from the Canadian High Arctic. Applied and Environmental Microbiology 76: 5773-5784. doi: 10.1128/aem.03094-09

Matschiavelli N, Oelgeschlager E, Cocchiararo B, Finke J, Rother M. (2012). Function and Regulation of Isoforms of Carbon Monoxide Dehydrogenase/Acetyl Coenzyme A Synthase in Methanosarcina

acetivorans. Journal of Bacteriology 194: 5377-5387. doi: 10.1128/jb.00881-12

McCalley CK, Woodcroft BJ, Hodgkins SB, Wehr RA, Kim E-H, Mondav R et al. (2014). Methane dynamics regulated by microbial community response to permafrost thaw. Nature 514: 478-+. doi: 10.1038/nature13798

McGuire AD, Macdonald RW, Schuur EAG, Harden JW, Kuhry P, Hayes DJ et al. (2010). The carbon budget of the northern cryosphere region. Current Opinion in Environmental Sustainability 2: 231-236. doi: 10.1016/j.cosust.2010.05.003

Mondav R, Woodcroft BJ, Kim E-H, McCalley CK, Hodgkins SB, Crill PM et al. (2014). Discovery of a novel methanogen prevalent in thawing permafrost. Nature Communications 5. doi: 10.1038/ncomms4212

Natali SM, Schuur EAG, Trucco C, Pries CEH, Crummer KG, Lopez AFB. (2011). Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra. Global Change Biology 17: 1394-1407. doi: 10.1111/j.1365-2486.2010.02303.x

Natali SM, Schuur EAG, Webb EE, Pries CEH, Crummer KG. (2014). Permafrost degradation stimulates carbon loss from experimentally warmed tundra. Ecology 95: 602-608. doi: 10.1890/13-0602.1

Orellana LH, Rodriguez-R LM, Higgins S, Chee-Sanford JC, Sanford RA, Ritalahti KM et al. (2014). Detecting Nitrous Oxide Reductase (nosZ) Genes in Soil Metagenomes: Method Development and Implications for the Nitrogen Cycle. mBio 5. doi: 10.1128/mBio.01193-14

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25: 1043-1055. doi: 10.1101/gr.186072.114

Peltier WR. (2004). Global glacial isostasy and the surface of the ice-age earth: The ice-5G (VM2) model and grace. Annual Review of Earth and Planetary Sciences 32: 111-149. doi: 10.1146/annurev.earth.32.082503.144359

Penton CR, StLouis D, Cole JR, Luo Y, Wu L, Schuur EG, Zhou J, Tiedje JM. (2013). Fungal diversity in permafrost and tallgrass prairie soils under experimental warming conditions. Applied and environmental microbiology. Nov 15;79(22):7063-72. doi: 10.1128/AEM.01702-13

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. Plos One 5. doi: 10.1371/journal.pone.0009490

Price PB, Sowers T. (2004). Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. Proceedings of the National Academy of Sciences of the United States of America. Mar 30;101(13):4631-6. doi: 10.1073/pnas.0400522101

Pries CEH, Schuur EAG, Crummer KG. (2012). Holocene Carbon Stocks and Carbon Accumulation Rates Altered in Soils Undergoing Permafrost Thaw. Ecosystems 15: 162-173. doi: 10.1007/s10021-011-9500-4

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590-D596. doi: 10.1093/nar/gks1219

Rho M, Tang H, Ye Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Research 38. doi: 10.1093/nar/gkq747

Roberts DL, Zhao SY, Doukov T, Ragsdale SW. (1994). The reductive acetyl coenzyme A pathway: sequence and heterologous expression of active methyltetrahydrofolate-corrinoid/iron-sulfur protein methyltransferase from Clostridium thermoaceticum. Journal of Bacteriology 176: 6127-6130.

Rodriguez-R LM, Konstantinidis KT. (2014). Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics 30: 629-635. doi: 10.1093/bioinformatics/btt584

Salles JF, Le Roux X, Poly F. (2012). Relating phylogenetic and functional diversity among denitrifiers and quantifying their capacity to predict community functioning. Frontiers in Microbiology 3. doi: 10.3389/fmicb.2012.00209

Schuur EAG, Abbott B. (2011). High risk of permafrost thaw. Nature 480: 32-33.

Sekiguchi Y, Muramatsu M, Imachi H, Narihiro T, Ohashi A, Harada H et al. (2008). Thermodesulfovibrio aggregans sp nov and Thermodesulfovibrio thiophilus sp nov., anaerobic, thermophilic, sulfate-reducing bacteria isolated from thermophilic methanogenic sludge, and emended description of the genus Thermodesulfovibrio. International Journal of Systematic and Evolutionary Microbiology 58: 2541-2548. doi: 10.1099/ijs.0.2008/000893-0

Shaver GR, Chapin FS. (1980). Response to fertilization by various plant-growth forms in an Alaskan tundra – nutrient accumulation and growth. Ecology 61: 662-675. doi: 10.2307/1937432

Singh BK, Quince C, Macdonald CA, Khachane A, Thomas N, Abu Al-Soud W et al. (2014). Loss of microbial diversity in soils is coincident with reductions in some specialized functions. Environmental Microbiology 16: 2408-2420. doi: 10.1111/1462-2920.12353

Stamatakis A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690. doi: 10.1093/bioinformatics/btl446

Tarnocai C, Canadell JG, Schuur EAG, Kuhry P, Mazhitova G, Zimov S. (2009). Soil organic carbon pools in the northern circumpolar permafrost region. Global Biogeochemical Cycles 23. doi: 10.1029/2008gb003327

Taş N, Prestat E, McFarland JW, Wickland KP, Knight R, Berhe AA et al. (2014). Impact of fire on active layer and permafrost microbial communities and

metagenomes in an upland Alaskan boreal forest. Isme Journal 8: 1904-1919. doi: 10.1038/ismej.2014.36

Torsvik V, Goksoyr J, Daae FL. (1990). High diversity in DNA of soil Bacteria. Applied and Environmental Microbiology 56: 782-787.

Tveit AT, Urich T, Svenning MM. (2014). Metatranscriptomic Analysis of Arctic Peat Soil Microbiota. Applied and Environmental Microbiology 80: 5761-5772. doi: 10.1128/aem.01030-14

UniProt C. (2015). UniProt: a hub for protein information. Nucleic acids research 43: D204-212. doi: 10.1093/nar/gku989

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73: 5261-5267. doi: 10.1128/aem.00062-07

Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences of the United States of America 95: 6578-6583. doi: 10.1073/pnas.95.12.6578

Wild B, Schnecker J, Alves RJ, Barsukov P, Bárta J, Čapek P, et al. (2014). Input of easily available organic C and N stimulates microbial decomposition of soil organic matter in arctic permafrost soil. Soil Biology and Biochemistry. Aug 31;75:143-51. doi:10.1016/j.soilbio.2014.04.014

Xue K, Yuan MM, Shi ZJ, Qin Y, Deng Y, Cheng L et al. (2015). Rapid microbial feedbacks reveal vulnerability of tundra soil carbon to climate warming. Nature Climate Change. doi: 10.1038/nclimate2940 [Epub ahead of print]

Yergeau E, Hogues H, Whyte LG, Greer CW. (2010). The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. Isme Journal 4: 1206-1214. doi: 10.1038/ismej.2010.41

Zerbino DR, Birney E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18: 821-829. doi: 10.1101/gr.074492.107

Zhang J, Kobert K, Flouri T, Stamatakis A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30: 614-620. doi: 10.1093/bioinformatics/btt593

Zhang X, Johnston ER, Liu W, Li L, Han X. (2016). Environmental changes affect the assembly of soil bacterial community primarily by mediating stochastic processes. Global change biology. Jan 1;22(1):198-207. doi: 10.1111/gcb.13080

Zhou J, Xue K, Xie J, Deng Y, Wu L, Cheng X et al. (2012). Microbial mediation of carbon-cycle feedbacks to climate warming. Nature Climate Change 2: 106-110. doi: 10.1038/nclimate1331

Zhu W, Lomsadze A, Borodovsky M. (2010). Ab initio gene identification in metagenomic sequences. Nucleic Acids Research 38. doi: 10.1093/nar/gkq275

CHAPTER 5: CONTRASTING RESPONSES OF TUNDRA SOIL MICROBIAL COMMUNITIES AT TWO CRITICAL DEPTHS AFTER FIVE YEARS OF EXPERIMENTAL WARMING

Eric R Johnston, Janet K Hatt, Kai Xue, Mengting Maggie Yuan, Shi Zhou, Zhili He, Liyou Wu, Yiqi Luo, Edward AG Schuur, James M Tiedje, Jizhong Zhou, & Konstantinos T Konstantinidis

5.1. Abstract

Northern-latitude tundra soils harbor substantial carbon stocks that are highly susceptible to microbial degradation with rising global temperatures. In this study, Alaskan tundra soil communities subjected to five years of experimental warming *in situ* were evaluated with shotgun-metagenomic sequencing at two depths: 15-25 cm (active layer) and 45-55 cm (recently-thawed permafrost boundary layer). Warming increased soil temperature evenly across the depth profile, but resulted in a considerably larger increase to annual thaw duration at 45-55 cm (+74.3%) relative to 15-25 cm (+7.7%). Among many shifts caused by warming at 45-55 cm, genes for methanogenesis and respiratory nitrate and sulfate reduction increased in relative abundance, whereas cytochrome c oxidase genes decreased from interactive effects of increased temperature and thaw depth. Increased methanogenesis potential coincided with a 3-fold increase in the dominant archaeal clade *Methanosarcinales*, for which genomes were recovered bioinformatically and found to encode genes for methanogenesis from

acetate, CO_2, and methanol. In contrast to small or insignificant differences after 1.5 years of experimentation, five years of warming resulted in significant changes to the abundances and composition of functional traits involved in carbohydrate metabolism at 15-25 cm. These shifts were more attributable to the direct effects of elevated temperature rather than indirect effects, such as a 25.2% increase in aboveground biomass. Collectively, our results demonstrated the speed of tundra community responses to moderate warming, greater susceptibility of profiles at or adjacent to the permafrost-layer boundary, and highly contrasting modes of response between either critical depth profile.

5.2. Introduction

Representing only ~16% of Earth's terrestrial surface, northern-latitude permafrost soils and their overlaying active layers harbor an estimated 1,672 to 1,832 Pg of carbon (C) in the form of soil organic C (SOC), which accounts for ~50% of the global SOC reservoir (Tarnocai et al., 2009; Ciais et al., 2013; Mu et al., 2015). This large C stock has accumulated and been preserved for thousands of years, primarily due to low temperatures and frozen conditions, which constrain microbial metabolisms involved in complete SOC mineralization (Schuur et al., 2008; Hicks Pries et al., 2012). Elevated atmospheric greenhouse gas concentrations are increasing global temperatures, and northern-latitude areas are experiencing a rate of warming that is more than twice the global average (IPCC 2013). As a result, regionally widespread and ongoing permafrost thaw is being observed (Jorgenson et al., 2001; Lawrence and Slater 2005; Osterkamp 2007; Romanovsky et al., 2010), and it has been estimated that permafrost could recede further by 30-70% by the end of the 21st century (Schuur and Abbott 2011; Lawrence et al., 2012). The resulting alleviation of prior abiotic

constraints is expected to stimulate microbial metabolisms involved in the release of greenhouse gases, particularly CO_2 and CH_4, threatening a potentially large positive feedback to climate warming (Heimann and Reichstein 2008; Schuur et al., 2013; Abott et al., 2016). Yet, despite the importance of delineating microbial response mechanisms for improving future climate change predictions, efforts to do so are hampered by the enormous complexity of soil microbial communities, a high degree of compositional heterogeneity *in situ* and across physicochemical gradients, and large knowledge gaps regarding their metabolic functions and other ecological attributes.

It is expected that deeper tundra soil, which can be altogether frozen (e.g., permafrost) and buffered by seasonal and diurnal temperature extremes, will be comparatively more vulnerable to temperature rise. For instance, previous investigations of lab-incubated permafrost soils under elevated temperatures reported observable shifts in community structure and functioning over a duration of weeks (Mackelprang et al., 2011; Coolen and Orsi 2015). Increasing temperatures and active layer thickness (i.e., an extension of maximum annual thaw depth) also affects other components of the soil, such as redox conditions, water availability, nutrient cycling, and aboveground vegetation (Schuur et al., 2009; Hinzman et al., 2013, Lipson et al., 2015; Keuper et al., 2012; Penton et al., 2016). These ecological alterations, in turn, can result in additional, indirect effects of warming to soil communities. For instance, gradual shifts in vegetative communities can alter soil physical conditions, establish new plant-microbe symbioses, promote oxygen and/or or methane transit, and change the amount and type of plant material inputted to soil (Bais et al., 2006; Grogan et al., 2001; Ström et al. 2005; Hicks Pries et al., 2013). Previous investigation of a temperature-limited steppe ecosystem found plant and soil communities shifting

in concert under conditions simulating future climate change (Zhang et al., 2017). A combination of soil warming and redox conditions favorable for energy-yielding activities can result in a more rapid release of CO_2 from tundra soils (Elbering et al., 2013; Lee et al., 2012). Under less favorable redox conditions, microbes can degrade SOC by gradually using electron acceptors with declining redox potentials, and conditions can eventually become suitable for methanogenesis (Mackelprang et al., 2016). This can result in lower C loss compared to oxygenated conditions, and yet have greater feedback to climate due to larger CH_4:CO_2 of emissions (Lee et al., 2012).

Hence, the microbial and environmental features that serve as potential candidates for ecosystem-level evaluation of soil C release are vast and interconnected. Recent advances in next-generation sequencing now allow for near-complete complete representation of the DNA materials recovered from tundra soils, due in part to lower diversity in relation to temperate grassland and agriculture habitats (Johnston et al., 2016). This genetic information can then be used to investigate numerous microbial mechanisms that potentially interact with other ecological factors under transitioning climate conditions. However, due to the recent nature of these advancements and other valuable ‘omics approaches, investigations of complex soils at the whole-community level are only beginning to emerge. There remain large gaps in our understanding of tundra microorganisms involved in CO_2 and CH_4 generation and the degree to which these activities extend from and/or contribute to processes occurring at regional and global scales. An improved understanding of the microbial clades and functions involved could help predict greenhouse gas fluxes and lead to an improved assessment of ecosystem-scale responses to climate warming (McCalley et al., 2014).

To advance these topics, the Carbon in Permafrost Experimental Heating Research (CiPEHR) site was established in a discontinuous permafrost region in Interior Alaska in 2008, and represents the first experiment to degrade surface permafrost without delaying spring soil melt (Natali et al 2011; Natali et al., 2012). In warmed plots, soil is insulated from frigid conditions during the winter months with snow fences that increase aboveground winter snow accumulation. This results in soil temperatures that are elevated year-round and also in the warming of deep soil, a manipulation not typically achieved with previous tundra warming experiments. Previous research at this site demonstrated rapid physiological responses after just two winters of experimentation. However, community-wide shifts in microbial structure were less discernible at this early stage, perhaps due to a lag between physiological response and corresponding shifts at the DNA level (e.g., cellular proliferation of responsive taxa is needed to observe differences in community structure) (Xue et al., 2016; Johnston et al., 2016). In this study, shotgun-metagenomic sequencing was used to assess responses of soil community structure to 5 years of experimental warming at two depths: 15-25 cm (pre-existing active layer) and 45-55 cm (the permafrost boundary layer at the initiation of experiment). The specific objectives were to (i) compare the relative sensitivities of microbial communities at either depth to elevated temperatures, (ii) identify and contrast the modes of response between depths, and (iii) assess differences in the direct and indirect (e.g., thaw, moisture, plant community, etc.) effects of warming.

5.3. Materials and methods

5.3.1. Study site description and sample collection

The Carbon in Permafrost Experimental Heating Research (CiPEHR) site was established at a moist acidic tundra area in September 2008 at an area in Interior Alaska near Denali National Park in the Eight Mile Lake region (63º52'59''N, 149º13'32''W). The experimental plots were located in the discontinuous permafrost region where permafrost thaw has been observed in the past several decades. Experimental design and site description were described in detail previously (Natali et al., 2011). Generally, three experimental blocks were located approximately 100 m away from each other. In each block, two snow fences were constructed about 5 m apart in the winter. The winter warming treatment plots were located 5 m back from the leeward side of the snow fences, while the paired control plots were at the windward side of snow fences. Soil temperature was increased in the winter warming plots due to thicker snow cover on soil surface and lower wind strength. The snow fences were removed in the spring before the snow melting to uniform hydraulic condition in both winter warming and control treatments.

From 1976 to 2009, mean monthly temperature in the field ranged from –16 °C in December to 15 °C in July, with an annual mean temperature of –1.0 °C. The average annual precipitation was 378 mm. Only C_3 plant species were observed in this area. Dominant species include *Eriophorum vaginatum*, *Vaccinium uliginosum*, some other vascular species, nonvascular feather moss and lichen. In the experimental plots, the upper 45-65 cm soil was rich in organic C materials and below was mineral soil with a mixture of glacial till and windblown loess. The active layer depth was about 50 cm at initiation of the

experiment, but has since expanded to lower depths. Twelve soil cores, six from each treatment, were taken using electric drills in destructive sampling plots at the six snow fences in the beginning of 2013 growing season (May), four and half year after the start of the winter warming manipulation.

5.3.2. Site monitoring and characterization of environmental indices

Thaw depth was measured weekly during the growing season (May to September) using a metal depth probe. The thaw depth data presented in this study were the average values for the 2010 growing season. Constantan–copper thermocouples and CR1000 data loggers (Campbell Scientific) were used to measure and record soil temperature and moisture content at each depth every half hour in flux bases installed in each plot. Volumetric water content from the soil surface to 15 cm depth was measured using site-calibrated Campbell CS616 water content reflectometer probes Soil moisture data presented in this study were averaged over the growing season.

To prepare soils for microbial and chemical analyses, visible roots and stones were removed by metal forceps. To measure soil C and nitrogen (N), soil samples (5 g) were dried at 70 °C until constant weight, ground to powder, encapsulated in silver foil and fumigated with HCl for 24 h at room temperature to remove soil inorganic C (carbonates). Soil C and N concentrations were analyzed in the Colorado Plateau Stable Isotope Laboratory at the Northern Arizona University on a DELTA V Advantage isotope ratio mass spectrometer (Thermo Fisher Scientific), configured through a Finnigan CONFLO III (Thermo Fisher Scientific) and using a Carlo Erba NC2100 elemental analyser (CE Elantech). The total organic C (TOC) and soil N content of each sample were calculated as the percentage mass of C or N.

5.3.3. Soil DNA isolation, metagenome library preparation, and sequencing

Soil DNA was extracted using a PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) according to manufacturer's protocol. DNA quality was assessed based on spectrometry absorbance at wavelengths of 230nm, 260nm and 280nm detected by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., now NanoDrop Products by Thermo Fisher Scientific). The absorbance ratios of 260/280 nm were around 1.8, and of 260/230 nm were larger than 1.7. Quantification of DNA isolated from soil was performed using a Qubit 2.0 Fluorometer and the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Dual-indexed DNA sequencing libraries were prepared using the Illumina Nextera XT DNA library prep kit according to manufacturer's instruction except that the protocol was terminated after isolation of cleaned double-stranded libraries. Prepared library DNA concentrations were determined with a Qubit HS DNA assay and libraries were run on a High Sensitivity DNA chip using the Bioanalyzer 2100 instrument (Agilent) to determine library average insert sizes. DNA libraries were sequenced at the Georgia Institute of Technology High Throughput DNA Sequencing Core on an Illumina HiSeq 2500 instrument in the rapid run mode for 300 cycles (150 bp, paired-end mode). Adapter trimming and demultiplexing of sequences (assignment of sequences to samples) were carried out by the Illumina 2500 instrument.

5.3.4. Bioinformatics data analyses

Metagenomic paired-end reads were merged using PEAR (Zhang et al., 2014) (options: -p 0.001). All merged and non-merged reads were then quality-trimmed with the SolexaQA package (Cox et al., 2010) (options: -h 17; ≥ 98% accuracy per nucleotide position). Trimmed sequences used downstream for functional annotation or taxonomic assignment were truncated to 150 bp to avoid read-length biases.

For the assessment of taxonomic composition, 16S and 23S rRNA gene fragments were first recovered from metagenomes using SortMeRNA (Kopylova et al., 2012). The relative proportion of sequences matching to bacterial vs. archaeal rRNA gene sequences from SortMeRNA was used to estimate the relative abundances of either domain. For this, abundances were first derived for 16S and 23S rRNA genes independently, and then averaged. The relative abundances of bacterial phyla were determined with Parallel-META 2.0 (Su et al., 2014), using SortMeRNA output sequences as input. Archaeal phylum-level classification summaries were determined by aligning SortMeRNA output sequences against the SILVA SSU database (v132) (Quast *et al.*, 2013) using BLASTN alignment (BLAST+ version 2.2.28; options: -word_size 18) (Camacho et al., 2009). Matches to archaeal sequences with ≥ 80% nucleotide identity and ≥ 100 bp alignment length were retained. To calculate the abundances of archaeal phyla in each metagenome as a proportion of the whole prokaryotic community, the number of sequences matching to each phylum were first normalized by the total number of sequences matching to Archaeal SILVA SSU references and then multiplied by the relative abundance of Archaea (as determined above). In order to assess community phylogenetic structure, SortMeRNA output sequences were used with the Qiime software package

(QIIME 1) (Caporaso et al., 2010) for closed-reference OTU picking with UCLUST (options: -m uclust_ref -C -z) (Edgar 2010) against the Greengenes database dereplicated at 99% nucleotide identity (DeSantis et al., 2006). Resulting OTUs with < 3 counts in all 24 sample metagenomes were removed from the OTU table to reduce noise. Pre-aligned Greengenes reference sequences represented in the OTU table were then used to construct a phylogenetic tree using FastTree (Price et al., 2010). The resulting phylogenetic tree and OTU count table was then used to generate a weighed UniFrac distance matrix summarizing pairwise phylogenetic distances between communities.

Short nucleotide sequences were searched against the Swiss-Prot (downloaded on 11-27-2016) (UniProt Consortium 2015) and CAZy (Carbohydrate-Active enZYmes; downloaded on 07-15-2016) (Cantarel et al., 2008) reference databases using DIAMOND BLASTX alignment (options: -k 1 -e 1E-5 --sensitive) (Buchfink et al., 2015). Matches to Swiss-Prot or CAZy reference sequences with a bit score $\geq$ 55 were retained for further analysis. Independent count matrices were made to summarize the results of alignment against either database. A reference data file associating Kegg Orthology functions (KO terms) (Kanehisa et al., 2016) to Swiss-Prot database entries was obtained from http://www.uniprot.org/downloads and was used to convert Swiss-Prot annotation counts into a more consolidated count matrix of KO terms. Swiss-Prot annotations were also consolidated into Gene Ontology (GO) categories as described previously in Johnston et al., 2016 (see Chapter 4). Annotations to CAZy reference sequences were consolidated into count matrix of structurally-related catalytic and carbohydrate-binding families. They were also consolidated further into the six broad functional modules covered in the CAZy database, including Glycoside Hydrolases (GH), GlycosylTransferases (GT),

Auxiliary Activities (AA), Carbohydrate Binding (CBM), Carbohydrate Esterases (CE), and Polysaccharide Lyases (PL). For each sample metagenome, the abundances of summarized CAZy definitions, as proportions of all functional genes, were determined by normalizing the number of annotations matching to each broad definition by the number of annotations matching to the more functionally-comprehensive SwissProt database as well as the total number of sequences originally used as query for functional annotation. For the summarized count matrices of KO terms or CAZy enzyme families (not summarized into modules), compositional dissimilarity between samples was calculated with Bray-Curtis and abundance weighted Jaccard distance metrics.

Merged and trimmed, non-truncated reads were assembled with IDBA-UD (Peng et al., 2012) (options: --mink 43 --maxk 123 --step 4 --min_contig 300). Resulting contigs $\geq$ 2 kbp were used with MetaBAT2 (options: --minCVSum 10) (Kang et al., 2015) and MaxBin2 (options: -min_contig_length 2000) (Wu et al., 2016) to bin assembled contigs for the recovery of microbial population genomes. Prior to binning, Bowtie 2 was used to align short-read sequences to assembled contigs (options: --very-fast) (Langmead and Salzberg 2012) and SAMtools was used to sort and convert SAM files to BAM format (Li et al., 2009). Sorted BAM files were then used to calculate the coverage (mean representation) of each contig in each sample metagenome. The quality of each resulting metagenome-assembled genome (MAG) was evaluated with CheckM v1.0.3 taxonomy workflow for Bacteria and Archaea, separately (Parks et al., 2015). The result from either evaluation (i.e., taxonomy worlkflow for Archaea or Bacteria) with the highest estimated completeness was retained for each MAG. Only MAGs with a quality score $\geq$ 60 were retained (from Parks et al., 2017; calculated as the estimated completeness minus five times the estimated

contamination). FastANI (Jain et al., 2017) was used to estimate the average nucleotide identity (ANI) between MaxBin2 and MetaBAT 2 generated MAGs. For redundant MAGs obtained independently from either binning method that matched at ≥ 95% ANI and were derived from the same assembled metagenome, the MAG version with the highest quality score was retained while the other was discarded. When these scores were identical, the CheckM lineage workflow was used to recalculate and compare scores. When quality scores remained identical after CheckM lineage workflow, the MAG version with the greatest overall size (in base-pairs) was retained. Protein-encoding genes from MAG contig sequences were predicted with Prodigal (Hyatt et al., 2010), and the resulting nucleotide sequences were searched against the Swiss-Prot database reference sequences using DIAMOND BLASTX alignment as described above, except that annotations were only retained if alignment/match length between the query and reference genes was ≥ 70% of the query or reference gene length, whichever was greater, and amino acid identity ≥ 40%. Select MAGs were also processed using the Microbial Genomes Atlas (MiGA) webserver for taxonomic assignment (Rodriguez-R et al., 2018).

5.3.5. Statistical analyses

With the R package nlme (Pinheiro et al., 2015), a linear mixed effects (LME) model treating experimental fence as a within-subjects factor was used to evaluate differences between experimental groups based on their environmental indices, their relative abundances of Bacteria, Archaea, and their phyla, the relative proportions of summarized CAZy definitions and broad Gene Ontology biological process terms. Phyla with an average relative abundance <0.1% were not considered for comparisons. ANOSIM and MRPP were used to evaluate the

significance of differences in warmed and control community functional and taxonomic/phylogenetic composition at either depth. Adonis was used to evaluate statistically significant associations between sample distances and pertinent environmental indices (e.g., thaw depth at sampling, annual thaw duration, etc.). Significance of correlations between metadata and certain community measures (metagenome annotations) was assessed using a two-tailed t-distribution. Tables with raw counts of functions (based on KO terms linked to Swiss-Prot reference annotations) were processed with the DESeq2 software package (Love et al., 2014) to identify significant, differentially abundant functions between sample depths (15-25 cm vs. 45-55 cm), treatments within a depth (where fence was treated as a within-subjects factor - e.g., using design = ~ Fence + Treatment), or using annual thaw duration as a continuous variable. To account for false discovery rate arising from multiple comparisons, p-values from DESeq2 analysis underwent Benjamini–Hochberg correction (Benjamini and Hochberg 1995). Results for functional annotations (i.e., KO terms, CAZy families, etc.) with an average relative abundance <2E-6 were typically discarded. The R package Superheat (Barter and Yu 2018) was used to illustrate the relative abundances of annotations in sample groups in conjunction with metadata and statistical results. The lm() function in R was used to perform multiple linear regression analysis to evaluate potential interactive effects between thaw depth and average summer 2012 temperature on the relative abundances of individual cytochrome C oxidase genes.

5.4. Results

5.4.1. Environmental Indices

Experimental warming by means of increased winter snow cover resulted in a year-round elevation of soil temperature. For instance, warming treatment plots were 1.02 and 1.11°C warmer than control plots at 15-25 cm and 45-55 cm soil depths, respectively, during the growing season prior to sampling (from May to September) (LME, $P < 0.05$) (Table 5-1). Experimental warming also increased annual thaw duration from 104.2 to 112.2 days at the 15-25 cm depth (7.7% increase) (non-significant with LME model, but $P < 0.05$ using paired t-test) and from 45.5 to 79.3 days in the 45-55 cm depth (74.4% increase) (LME, $P < 0.05$). The thaw depth at the time of sampling in May 2013 was 18.3 and 23.0 cm in control and experimentally warmed plots, respectively (LME, $P < 0.05$). The experimental warming treatments also increased mean aboveground plant biomass by 25.2%, increased volumetric water content slightly, and increased bulk density at the 45-55 cm depth profile (LME, $P < 0.05$). For aboveground plants, warming decreased lichen biomass by 41.8%, on average (paired t-test, $P < 0.05$) and increased the biomass of *Eriophorum vagninatum* and *Rhododendron tomentosum by* 81.2% and 23.3%, on average, respectively (paired t-test, $P < 0.05$). Paired t-test was used to compare certain ecological factors that did not involve a depth component (e.g., thaw depth at time of sampling).

Table 5-1: Summary of environmental and soil physiochemical measurements for the samples or experimental plots of the study. Values are given as the mean ± the standard error of the mean derived from six replicate measurements or soil. Superscript letters are used to distinguish sample groups that were significantly different (adj. p-value < 0.05), i.e., values with superscript letters differing from letters assigned to other values designates a statistically-significant difference between groups (i.e., treatment and depth). Statistical significance was determined using a linear mixed-effects model (where experimental fence was treated as a within-subjects factor), in conjunction with Tukey's HSD test.

	Annual Thaw Duration (days)	Growing Season '12 Temperate (°C)	Winter '12/'13 Temperature (°C)	May '13 Temperature (°C)
15-25 cm Control	104.2 ± 1.7 ^{c}	2.59 ± 0.29 ^{b}	-1.15 ± 0.16 ^{a}	-0.62 ± 0.1 ^{a}
15-25 cm Warmed	112.2 ± 1.8 ^{c}	3.61 ± 0.12 ^{c}	-0.66 ± 0.07 ^{bc}	-0.26 ± 0.04 ^{b}
45-55 cm Control	45.5 ± 4.6 ^{a}	1.08 ± 0.28 ^{a}	-0.86 ± 0.13 ^{ab}	-0.73 ± 0.13 ^{a}
45-55 cm Warmed	79.33 ± 5.55 ^{b}	2.19 ± 0.39 ^{b}	-0.42 ± 0.04 ^{c}	-0.3 ± 0.05 ^{b}

	Soil Moisture	Bulk Density	Total N (%)	Total C (%)
15-25 cm Control	0.74 ± 0.04 ^{b}	0.25 ± 0.04 ^{a}	1.58 ± 0.11 ^{b}	33.4 ± 3.4 ^{b}
15-25 cm Warmed	0.75 ± 0.03 ^{b}	0.23 ± 0.03 ^{a}	1.59 ± 0.07 ^{b}	34.6 ± 1.3 ^{b}
45-55 cm Control	0.39 ± 0.06 ^{a}	0.74 ± 0.17 ^{b}	0.65 ± 0.19 ^{a}	16.0 ± 4.1 ^{a}
45-55 cm Warmed	0.31 ± 0.03 ^{a}	1.19 ± 0.19 ^{c}	0.46 ± 0.11 ^{a}	12.4 ± 3.0 ^{a}

	Plant Biomass (g)	Thaw Depth (cm)	Water Table Depth (cm)	Volumetric Water Content (%)
Control Plots	582.3 ± 24.9 ^{a}	18.3 ± 0.5 ^{a}	22.96 ± 2.03 ^{a}	28.07 ± 0.14 ^{a}
Warmed Plots	728.9 ± 64.6 ^{b}	23.0 ± 1.6 ^{b}	22.29 ± 4.84 ^{a}	29.21 ± 0.55 ^{b}

5.4.2. Broad microbial community indices

On average, 8.1 Gbp (billion pase-pairs) of sequencing effort per sample was obtained for twelve soil communities representing the 15-25 cm soil depth and 5.8 Gbp per sample for twelve communities representing the 45-55 cm depth (i.e., 6 replicates per treatment x 2 treatments x 2 depths = 24 unique metagenomes; 166.8 Gbp total sequencing) (Table D-1). Following merging, quality trimming, and filtering, ≥80% of sequences were retained per metagenome and used for analyses. Using Nonpareil 3 (options: -T kmer -k 32 -X 100000) (Rodriguez-R et al., 2018), the average estimated coverage was 0.52 for 15-25 cm and 0.62 for 45-55 cm soil communities, respectively (Figure D-1). These coverages imply that, beyond the sequencing depth achieved, there is 52% and 62% likelihood that the additional sequence(s) obtained would be redundant with ones already observed. Thus, despite less sequencing data to represent the 45-55 cm communities, the comparatively lower diversity at this depth resulted in greater coverage by sequencing of these soil communities. The level of coverage obtained here (52-62%) should be appropriate for comparisons (e.g. low false negative rate), with un-sampled diversity presumably representing comparatively rarer taxa (Rodriguez-R and Konstantinidis 2014; Zhang et al., 2017). The amount of sequencing depth needed to achieve 95% average coverage was estimated to be 36.0 Gbp and 19.5 Gbp for 15-25 cm and 45-55 cm soil communities, respectively, which is lower than typical temperate soils and consistent with our previous study at an early time point (Johnston et al, 2016). The experimental warming treatment resulted in a near-significant increase in Nonpareil-derived α-diversity at the 45-55 cm depth (paired t-test, $P < 0.1$) (Figure D-2). MicrobeCensus, a tool that estimates the average genome size of microbial communities by evaluating the abundance of several single copy genes

(Nayfach and Pollard 2015), revealed smaller genomes, on average, of microbial populations comprising the 45-55 cm soil depth compared to those at 15-25 cm (average of 4.7 Mbp vs. 6.7 Mbp) (LME, $P < 0.001$) (Table D-2). Average genome sizes did not differ significantly between warmed and unwarmed soil communities at either depth (LME, $P > 0.1$).

5.4.3. Soil community taxonomic composition

The mean relative prokaryotic abundance of Archaea differed between soil depths; it was 1.69% and 2.64% in 15-25 cm and 45-55 cm soil depths, respectively (LME, $P < 0.05$) (Figure 5-1b). The archaeal community fraction at the 15-25 cm depth was mostly represented by phyla *Thaumarchaeota* followed by *Euryarchaota* at 66.2% and 24.0%, on average, respectively. For the 45-55 cm layer, archaeal 16S rRNA gene sequences were mostly represented by *Euryarchaeota* and *Crenarchaeota* at 88.1% and 9.1%, respectively. Warming had no significant effect on the total relative abundance of Archaea at the 15-25 cm depth (LME, $P > 0.1$), but increased by a mean of 39% at 45-55 cm (2.21% in control vs. 3.07% warmed soils; LME, $P < 0.05$). This increase was primarily driven by an increase in the abundance of archaeal phylum *Euryarchaeota* (1.90% vs. 2.73%), and more specifically, by order *Methanosarcinales*, which was ~3-fold more abundant in experimentally warmed plots at this soil depth (0.54% vs. 1.61% of the prokaryote community; LME, $P < 0.05$). Notably, a very strong correlation was observed between the relative abundance of *Methanosarcinales* and annual thaw duration (linear regression $R^2 = 0.804$, $P < 0.001$; squared roots $R^2 = 0.895$, $P < 0.001$) (Figure 5-1c).

Of the 18 bacterial phyla with mean relative abundances $\geq$ 0.1% across all 24 soil samples, most differed significantly between 15-25 cm and 45-55 cm soil

communities in relative abundance (14/18 in total, excluding *Thermotogae*, *Lentisphaerae*, OP10, and *Actinobacteria*; LME, $P < 0.05$) (Figure 5-1a). At 15-25 cm, warming only significantly increased the relative abundance of *Verrucomicrobia* (8.0% vs. 10.4% mean abundance in control and warmed samples, respectively; $P < 0.05$). In contrast, at 45-55 cm, experimental warming resulted in significant shifts ($P < 0.05$) to phyla *Proteobacteria* (23.4% vs. 30.4% mean abundance in control and warmed samples, respectively), *Actinobacteria* (31.0% vs. 20.6%), and OP11 (0.2% vs 0.8%), and near-significant shifts ($P < 0.1$) to *Bacteroidetes* (5.1% vs. 7.5%), OP10 (0.2% vs. 0.4%), and *Themotogae* (0.2% vs. 0%). Similarly, at 45-55 cm, experimental warming resulted in significant community-wide shifts in phylogenetic and taxonomic β-diversity represented by 16S-based weighted UniFrac distances (MRPP, $P < 0.05$) and abundance weighted Jaccard distances of phylum-level community composition (ANOSIM and MRPP, $P < 0.05$) (Table 5-2) (Figure 5-2). There was also a significant relationship between β-diversity (derived from either UniFrac or Jaccard distances) and annual thaw duration (Adonis, $P < 0.05$) and between weighted UniFrac distances and the depth of thaw at time of sampling (Adonis, $P < 0.01$). There were no significant relationships between experimental treatment or thaw indices and 16S-based or phyla-based evaluations of community structure for samples representing the 15-25 cm depth, similar to observations of this depth made after just two winters of experimental warming at this site (Xue et al., 2016).

Table 5-2: The effects of warming and thaw depth or duration on microbial community functional and phylogenetic structure, for each depth. Permutation procedure (MRPP) and analysis of similarity (ANOVA) were used to test the effect of experimental warming, and permutational multivariate analysis of variance (Adonis) was used to test the effect of thaw depth (at time of sampling; in cm) and the annual duration of thaw (in days of the year) on each community metric. Numerical values represent probability scores (i.e., p-values) resulting from each test. P-values < 0.1 are bolded to highlight significant and near-significant results.

			Adonis		***ANOSIM***	***MRPP***
	Underlying Data	**Distance Metric**	**Thaw Time**	**Thaw Depth**	**Treatment**	**Treatment**
15-25 cm Soil Depth	16S-based OTUs	Weighted Unifrac	0.585	0.226	0.534	0.449
	Phyla Composition	Bray-Curtis	0.463	0.165	0.398	0.442
		Abund. Jaccard	0.455	0.479	0.915	0.747
	Kegg Ontology Terms	Bray-Curtis	0.416	0.761	0.894	0.769
		Abund. Jaccard	0.491	0.576	0.440	0.119
	CAZy Families	Bray-Curtis	0.232	0.441	0.543	0.346
		Abund. Jaccard	0.330	0.284	**0.083**	**0.047**
45-55 cm Soil Depth	16S-based OTUs	Weighted Unifrac	**0.010**	**0.003**	0.204	**0.010**
	Phyla Composition	Bray-Curtis	**0.018**	**0.002**	0.131	**0.087**
		Abund. Jaccard	**0.045**	**0.053**	**0.043**	**0.039**
	Kegg Ontology Terms	Bray-Curtis	**0.048**	**0.069**	**0.091**	0.139
		Abund. Jaccard	**0.010**	**0.006**	**0.099**	**0.068**
	CAZy Families	Bray-Curtis	0.170	0.152	**0.014**	0.142
		Abund. Jaccard	0.523	0.445	**0.095**	0.438

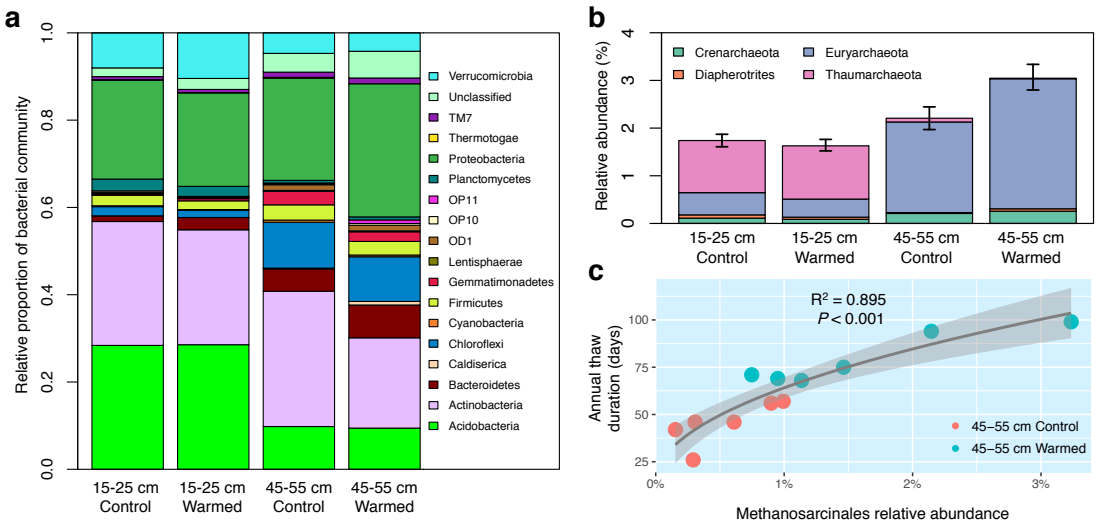


Figure 5-1: Taxonomic shifts as an effect of experimental warming. (a) Mean relative abundance of bacterial phyla for each depth x treatment combination. Underlying data is based on 16S rRNA gene-encoding fragments recovered from metagenomic datasets. Values represent the abundance of each bacterial phylum as a proportion of the total bacterial community. Only phyla with a mean relative abundance across all 24 datasets $\geq$ 0.1% are displayed. **(b) Mean relative abundances of archaeal phyla for each depth x treatment combination.** Underlying data is based on 16S rRNA gene-encoding fragments recovered from metagenomic datasets. Values represent the abundance (as a percentage) of each archaeal phylum relative to the total abundance of all recovered bacterial and archaeal 16S rRNA gene fragments (i.e., the total prokaryotic community). Error bars represent the mean ± the standard error of the mean (n = 6) for cumulative (total) relative *Archaea* abundance. **(c) Correlation between the relative abundance of *Methanosarcinales* and annual thaw duration (in days of the year) for 45-55 cm soils.** Linear regression was fitted for n = 12 points. Significance of the correlation coefficient was determined using a two-tailed Student's t-distribution (r = 0.946, df = 11).

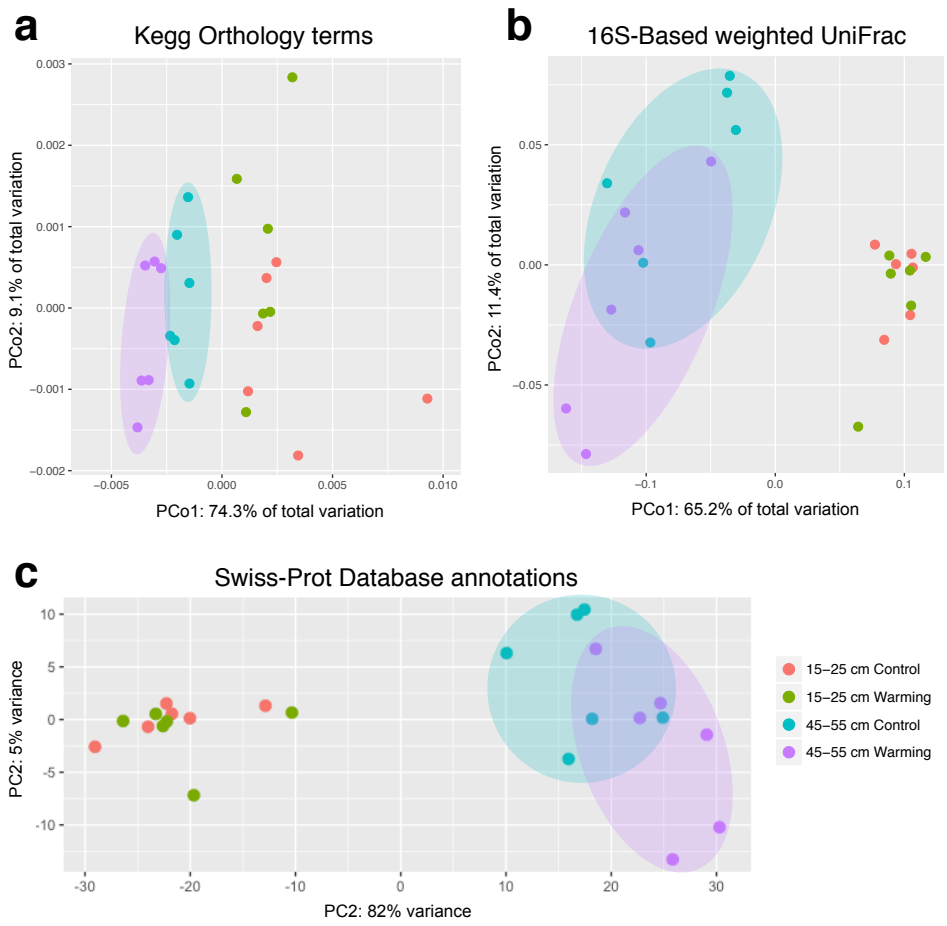


Figure 5-2: Functional and phylogenetic shifts as an effect of experimental warming. (a) PCoA plot of consolidated KEGG Ontology term annotations. Underlying data is based on abundance-weighted jaccard distance matrix derived from a Kegg Ontology term counts matrix. Abundances of Kegg Ontology terms for each sample were consolidated from Swiss-Prot database references annotations. **(b) PCoA plot of community phylogenetic composition.** Underlying data is a weighted unifrac distance matrix of 16S rRNA gene-encoding fragments recovered with Parallel-META and processed in the QIIME software package as described in the Methods section. (**c**) **PCA plot of Swiss-Prot gene annotations.** Underlying data is based on a gene count matrix

(consolidated from Swiss-Prot database references), which underwent variance-stabilizing transformation using the DESeq2 package. The 24 samples shown in the 2D plane spanned by their first two principal components.

5.4.4. Soil community functional composition

In addition to the observed shifts in microbial community phylogenetic and taxonomic structure for 45-55 cm metagenomes, experimental warming resulted in significant or near-significant shifts to community functional structure (summarized as CAZy families or KO terms) (ANOSIM, $P < 0.05$ or $P < 0.1$) (Table 5-2; Figure 5-2). Also for this depth, β-diversity based on KO terms was significantly associated with annual thaw duration and measured thaw depth at time of sampling (Adonis, $P < 0.05$). For 15-25 cm soils, warming only significantly altered community β-diversity reflecting the composition of genes involved in carbohydrate metabolism and binding (i.e., CAZy protein families as underlying data) (MRPP, $P < 0.05$; ANOSIM, $P < 0.1$), and no significant relationships between functional structure and either thaw index for this depth were identified.

Consistent with shifts in β-diversity based on CAZy protein families observed for 15-25 cm communities, experimental warming increased relative proportion of functional genes matching to CAZy reference sequences by 7.9% (relative to the number of Swiss-Prot database annotations; % change and significance was similar when normalizing by query sequences used for alignment instead) (LME, $P < 0.05$) (Figure 5-3a). This included a 12.0% increase in glycoside hydrolase genes, a 2.6% increase in glucosyltransferase genes, a 8.1% increase in carbohydrate binding genes, a 8.9% increase in carbohydrate

esterase gene, and a 20.8% increase in polysaccharide lyase genes (LME, $P < 0.05$) (Figure 5-3ab). While warming also shifted community β-diversity of CAZy families at the 45-55 cm depth, there was only a significant 3.1% increase in the abundance of glucosyltransferase genes at this depth (LME, $P < 0.05$). These results were generally consistent with Gene Ontology process categories involving the catabolism of various organic matter substrates (e.g., “cellulose catabolic process”) (Figure D-3). Of the 25 categories based on these definitions that differed by $\geq$ 10% between control and warmed 15-25 cm soils, 22/25 were more abundant in experimentally warmed metagenomes, with four having adj. $P < 0.05$ with LME, (but none were significant when evaluated with DESeq2). In contrast, significant differences in the relative proportion of functional genes matching to all CAZy references or summarized modules were not found between warmed and unwarmed 15-25 cm soils from the 1.5-year collection (Figure D-4).

Warming of the 15-25 cm soil layer did not result in any KO terms that changed in abundance with an adj. $P < 0.05$ (using DESeq2). This result was observed whether the statistical design included fence as a within-subjects factor and without it. However, of the 3,274 KO terms with abundances adequate for p-value assignment in DESeq2, 162 KO terms differed significantly (adj. $P < 0.05$) and another 153 differed near-significantly ($P < 0.1$) between control and warmed soil communities samples at 45-55 cm. Also for the 45-55 cm depth, the abundances of 490 KO terms increased or decreased significantly across a gradient of annual thaw duration (i.e., when thaw duration was used as the ‘condition’ for DESeq2’s statistical design), and another 301 KO terms displayed a near-significant change with thaw duration. These shifts included an increase in many functions involved in methanogenesis, many of which differed significantly

or near-significant (adj. $P < 0.05$ or < 0.1) between control and experimentally warmed communities (Figure 5-4a). This evaluation was primarily based on KEGG modules 'M00567', 'M00357', 'M00356', 'M00563', corresponding to methanogenesis from CO_2, acetate, methanol, and methylamines (Kanehisa et al., 2016) (Figure 5-4a). Genes encoding acetate kinase and phosphate acetyltransferase were not used in this evaluation due to their participation in other metabolic pathways (e.g., glycolysis). None of the 40 KO terms used to assess methanogenesis potential decreased with warming or demonstrated a negative relationship with annual thaw duration (i.e., none with a log2-fold-change < 0).

One of the largest and consistently significant responses was observed for KO terms specific to methanogenesis from acetate, which increased on average by 170% vs. a 54.5% increase in *mcrA*, an essential marker gene possessed by all methanogens. There were also strong correlations between annual thaw duration and the summarized mean abundance of KO terms specific to methanogenesis from acetate ($R^2 = 0.796$; unpaired Student's t-distribution, $P < 0.01$) or when using the mean abundance derived from all 40 of these KO terms ($R^2 = 0.723$; $P < 0.01$). These correlations with methanogenesis genes were stronger than those obtained using other environmental indices (e.g., moisture, water table depth at time of sampling, volumetric water content, soil C%, etc.). There were also strong correlations between the abundance of archaeal order *Methanosarcinales* (see previous section) and KO terms for methanogenesis from acetate ($R^2 = 0.93$; $P < 0.01$) or the mean abundances of all methanogenesis functions ($R^2 = 0.81$; $P < 0.01$). Warming or extended annual thaw of 45-55 cm soils also increased the relative abundances of genes involved in dissimilatory sulfate reduction (*dsrA*, *dsrB*, *aprB;* DESeq2, adj. $P < 0.05$ with

thaw and/or treatment) (Figure 5-4c) and dissimilatory nitrate reduction to ammonium, (*nirB*, *nrfA*, and *nrfH;* DESeq2, adj. $P < 0.05$ or 0.1 with thaw and/or treatment) (Figure 5-4b), but resulted in no significant changes in denitrification genes *nirk*, *norB*, and *nosZ*. The opposite trend was observed for KO terms specific to assimilatory nitrate reduction (*nirA*) and assimilatory sulfate reduction (*sir*, *cysI*) (Figure 5-4bc). DESeq2 revealed statistically significant decreases in the relative abundances of cytochrome c oxidase genes with average 2012 growing season temperature, even when controlling for thaw indices. Multiple linear regression analysis revealed that average summer 2012 temperature and thaw depth had significant, independent (controlling for the other variable) as well as additive negative associations with the abundances of cyctochrome c oxidase genes (e.g., cytochrome c oxidase subunit 1 and 2; multiple $R^2 > 0.8$; $P < 0.001$) (Figure D-5).

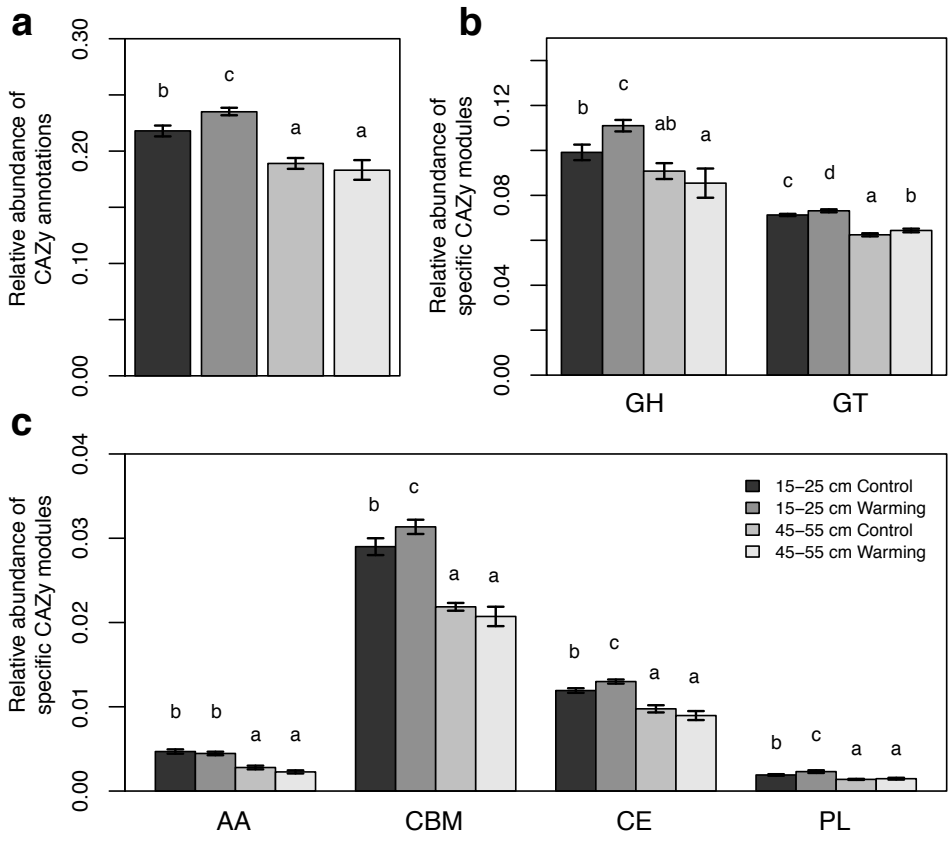


Figure 5-3: Shift in carbohydrate utilization (CAZy) genes as an effect of of experimental warming. Relative abundances of **(a)** all CAZy annotations, and CAZy modules **(b)** Glycoside Hydrolases (GH), GlycosylTransferases (GT), **(c)** Auxiliary Activities (AA), Carbohydrate Binding (CBM), Carbohydrate Esterases (CE), and Polysaccharide Lyases (PL). Underlying values represent the mean relative abundance of each category for each sample group (i.e., treatment x depth). Letters are to distinguish sample groups that were significantly different (adj. $P < 0.05$), i.e., values with letters differing from letters assigned to other groups designates a statistically-significant difference between groups. Significance was determined using a linear mixed-effects model (where experimental fence was treated as a within-subjects factor), in conjunction with Tukey's HSD test.

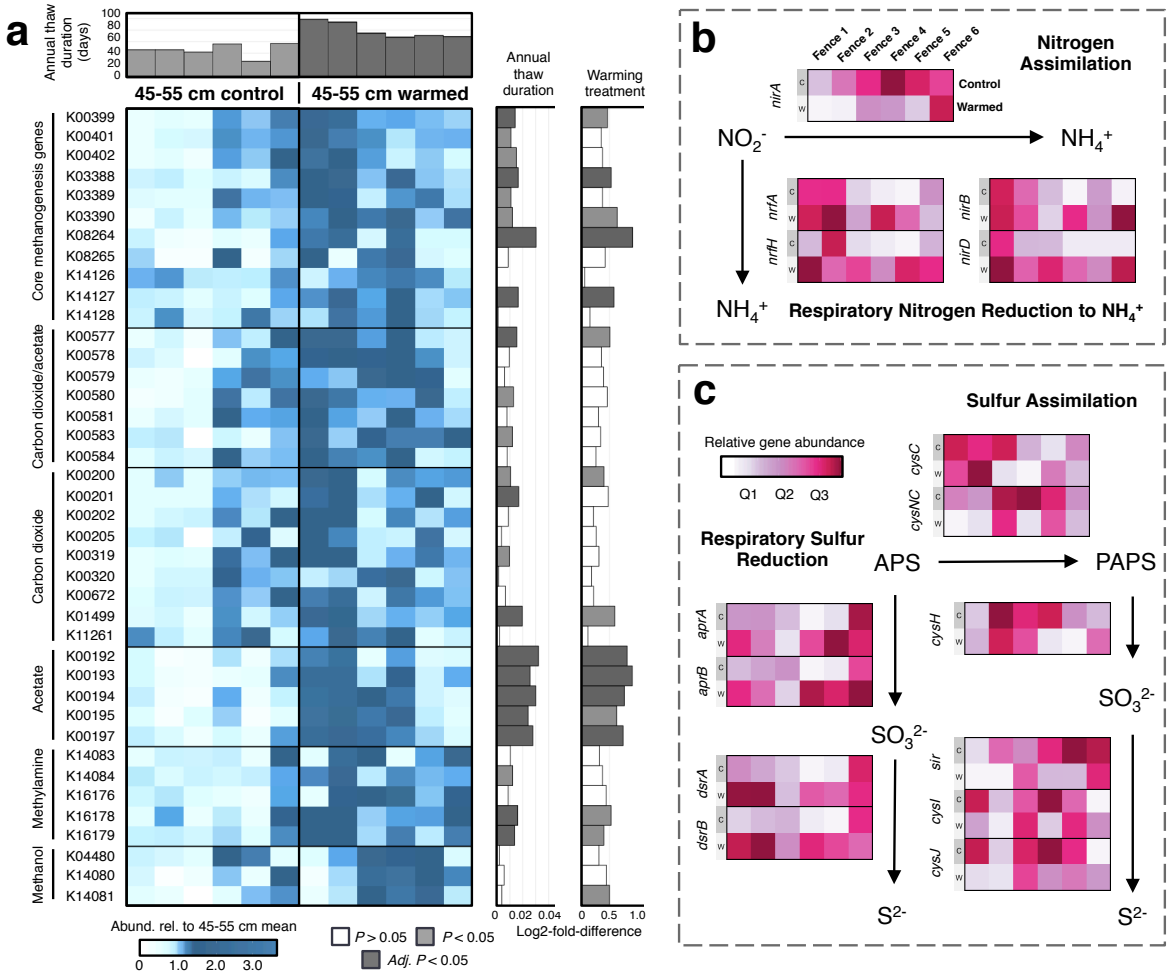


Figure 5-4: Shifts in microbial energy-generating pathways as an effect of experimental warming. (a) Heatmap showing the relative abundances of genes involved in methanogenesis for 45-55 cm depth soil metagenomes. Kegg Orthology (KO) terms under the 'Core methanogenesis genes' category represent those used by all methanogens. Other categories and KO terms refer to pathways involved in the usage of different electron acceptors to sustain methanogenesis, which are variable between methanogenic taxa. To emphasize differences between datasets, the values for each gene were normalized by the mean relative abundance of all twelve 45-55 cm soil metagenomes (by rows). Sub-plots on the right of the heatmap represent the log2-fold-difference

calculated for each gene, and are colored by whether differences between control and experimentally warmed soils were statitistical significant (right) or if there was a significant association between gene relative abundance and annual thaw duration (left) (dark grey = adj. p-value < 0.05; light grey = non-adj. p-value < 0.05; white = non-significant). Sub-plot on the top of the heatmap represents the annual thaw duration (i.e., days of the year soil was thawed) for each plot represented by the corresponding metagenomes (e.g., C1, C2, T1, etc.). **(b) Heatmaps showing the relative gene abundances for assimilatory and dissimilatory nitrate reductions. (c) Heatmaps showing the relative abundances for assimilatory and dissimilatory sulfate reduction.** Underlying data for (a) and (b) are communities respresenting six experimentally warmed and six control 45-55 cm soil samples.

5.4.5. Recovery of metagenome-assembled genomes (MAGs)

Assembly and binning led to the recovery of 173 metagenome-assembled genomes (MAGs) with a quality score $\geq$ 60. De-replication of highly similar MAGs based on FastANI values of $\geq$ 95% resulted in the consolidation of 74 non-redundant MAGs. This included the recovery of 11 (four were non-redundant) archaeal MAGs, all of which originated from 45-55 cm soil metagenomic datasets. Two non-redundant archaeal MAGs were identified as belonging to order *Methanosarcinales* and the other two were assigned to *Methanocellales* based on 16S and 23S rRNA gene assignment and taxonomic identification with MiGA. Each archaeal MAG was found to possess genes for methanogenesis, specifically for using CO_2, acetate, and in one case, methanol, as electron acceptors (Figure D-6). Among bacterial MAGs that shifted in abundance with

warming or along environmental gradients, there was a variable-yet-large 79% increase in the most dominant bacterial population (~1-2% of the community) in the 15-25 cm active layer with warming (paired t-test, $P < 0.1$) (Figure D-7). These results are consistent with smaller, non-significant shifts observed after 1.5 years of experimentation (Figure D-7). This *Acidobacteriaceae* population was previously found to be widespread throughout Alaskan tundra (mostly complete genome representation and ANI > 98%; over at least > 500 kilometers) and appears to encode diverse metabolisms for labile and recalcitrant organic matter degradation (Johnston et al., 2016).

5.5. Discussion

An improved understanding of tundra soil microbiota and their responses to warming conditions is critical for better modeling and predicting the magnitude of climate change, as a large degree of uncertainty persist regarding the vulnerability of tundra soil C and its potential to accelerate climate warming. The purpose of this investigation was to compare the functional and phylogenetic structure of tundra soil communities subjected to five years of accelerated warming *in situ* against unwarmed communities (i.e., those only subject to natural warming). Experimental manipulation began in September 2008 when the permafrost boundary layer was at a depth of ~50 cm (Natali et al., 2011), but active layer thickness has increased in subsequent years due to ongoing thaw observed in this region. Temperature increases lasted year-round and were moderately uniform between the depth profiles evaluated here (15-25 and 45-55 cm) (Table 5-1). Warming will undoubtedly result in changes to numerous important aspects of tundra ecology such as thaw and plant community factors, which can lead to indirect effects of warming on soil community functioning

(Sistla and Schimel 2013; Hinzman et al., 2013; Elberling et al., 2013). The magnitude and direction of their responses to these factors could vary considerably between depth profiles. Thus, warming of deep soil extending down to the permafrost boundary layer was a desirable outcome.

Experimental warming increased annual thaw duration by 7.7% (+8 days) (relative to unwarmed plots) at the 15-25 cm soil layer and, in stark contrast, by 74.4% (+33.8 days) at the 45-55 cm layer for the year preceding sample collection. Consistently, greater overall changes in community structure were observed for 44-55 cm soil communities. This includes a ~3-fold increase in *Methanosarcinales*, a methanogenic order and the most abundant archaeal clade observed in these soils, which were found capable of methanogenesis from CO_2, acetate, and methanol based on bioinformatically-recovered genome sequences (Figure D-6). There was also a close association between the abundance of this clade and annual thaw depth (R^2 = ~0.9, P < 0.001) (Figure 5-1c), which accompanied a comprehensive increase in the relative abundances of methanogenesis genes (Figure 5-4a). Methane oxidation genes were fairly low in relative abundance (e.g., < 1/100,000 annotations); those that were over the detection limit did not shift in response to warming at either depth. It is possible that methane was consumed at a depth not evaluated in this study (i.e., 25-45 or 0-15 cm) or by anaerobic methane oxidizing microbiota that remain poorly understood (Hallam et al., 2004; Thauer and Shima 2009). However, it has been demonstrated that CH_4 emissions at the CiPEHR site are increasing and were considerably greater from experimentally warmed plots (Natali et al., 2015). More recently, increased CH_4 emissions from this general area (the Eight Mile Lake Region) were found to account for more than half of C emissions based on global warming potential (Taylor et al., 2018).

For the last decade (from 2008-2017), the majority of tundra habitats have exhibited a 2-4.2 °C temperature increase above the 1950-1980 base period during cold seasons of the Northern Hemisphere (Nov - Apr) (Hansen et al., 2010; GISTEMP Team, 2018). Despite an increase in background temperature that has since thawed both control and experimentally warmed soil at the 45-55 cm depth that is altogether greater than the mean +0.45 to +1.11 °C temperature increase due to experimental warming (during the winter and growing seasons, respectively), a dramatic response in methanogenic taxa was observed nonetheless. Experimental warming also decreased the relative abundances of cytochrome oxidases that use O_2 as an electron acceptor. DESeq2 and multiple linear regression analyses results indicated that warming and thaw caused independent as well as additive declines in the abundances of these O_2-specific genes (Figure D-2). It is possible that accelerated warming causes a more thorough depletion of electron acceptors than would occur under gradual warming, due to a smaller period over which they can be replenished from oxygen penetration or water flow (containing dissolved electron acceptors) and result in a greater CH_4:CO_2 of emissions. A decline in redox conditions was also evidenced by the increased relative abundances of dissimilatory nitrate reduction to ammonium (DNRA) and dissimilatory sulfate reduction (Figure 5-4bc) genes. While we did not observe any changes in the relative abundances of denitrification genes, sulfide production (i.e., from dissimilatory sulfate reduction) has been found to inhibit denitrification and can also supply DNRA as an electron donor (An and Gardner 2002). Preference for DNRA has also previously been reported for conditions of high organic carbon relative to electron acceptor availability (Giblin et al., 2013). Accompanying declines in assimilatory nitrogen and sulfur reduction genes were also observed under experimentally warmed

conditions at 45-55 cm (Figure 5-4bc). These contrasting responses could be due to enhanced NH_4^+ production (from DNRA) and selection of traits for direct acquisition of NH_4^+ vs. those for NO_3^- acquisition and reduction to NH_4^+ for anabolism.

Warming also increased the biomass of *Eriophorum vaginatum* by 81%, a plant species that excretes organic acids such as acetate that can fuel CH_4 production, transit, and release (Ström et al. 2005). Another plant species also found to fuel methanogens, *Carex rostrata* (Strom et al., 2005), is common in this field location albeit with lower biomass. It is possible that a comparably greater response observed for genes specific to methanogenesis from acetate than other methanogenesis pathways (Figure 5-4a) is attributable to these types of plant-methanogen interactions. However, the associations between plant and methanogens resulting in greater CH_4 emission from tundra habitats remains elusive (Wang et al., 2008; Kao-Kniffin et al., 2010) and thus, is worthy of consideration in future studies.

At 15-25 cm, warming increased the total proportion of community functional genes involved in carbohydrate metabolism (Figure 5-3a), including glycoside hydrolases, glycosyltransferases, carbohydrate esterases, carbohydrate binding, and polysaccharide lyases (Figure 5-3bc). An early assessment of this depth profile after 1.5 years of experimental warming found a similar response using GeoChip that was undetected or less obvious with metagenomic analysis (Xue et al., 2016); thus, the response observed here is consistent with earlier observations, but also implies a gradual change over time (Figure D-4). Also consistent with the findings of Xue et al., 2016, a discernable shift in whole-community structure (e.g., β-diversity using 16S-based phylogeny or KO term abundances) was not observed. However, consolidation of CAZy

reference sequences into family definitions did reveal significance in the relationship between warming treatment and the relative composition (i.e., β-diversity) of these functions (Table 5-2). Nearly five years of experimental warming also increased plant biomass by 25.2% (paired t-test, $P < 0.05$), likely as a result of extended growing seasons and increased root colonization with enhanced thaw (Table 5-1), which could have contributed to the increase in genes related to carbohydrate metabolism. However, we observed a moderately-strong association between the fraction of genes involved in carbohydrate metabolism and the average 15-25 cm soil temperature during the same month for which soils were collected (May) ($R^2 = 0.521$; $P < 0.01$), as well as the average temperature of the preceding winter season ($R^2 = 0.478$; $P < 0.05$). For the six broad modules types represented in CAZy, carbohydrate esterases had the strongest relationship with May soil temperature at 15-25 cm ($R^2 = 0.644$; $P < 0.01$), followed by glycoside hydrolases ($R^2 = 0.532$; $P < 0.01$) and polysaccharide lyases ($R^2 = 0.453$; $P < 0.01$). These direct associations between the community fraction of genes involved in carbohydrate metabolism and soil temperate were much stronger than correlations from other broad environmental measure displayed in Table 5-1; this includes plant biomass, for which a significant relationship with metabolisms acting on carbohydrates was not observed ($R^2 = 0.179$ or less; $P > 0.1$). Mantel test used to evaluate associations between soil microbial community and plant community β-diversity (as Bray-Curtis distances of recorded plant biomass for each species) also did not indicate a significant relationship between communities. These results of stimulated functions involved in OC metabolism as a direct response to warming are somewhat surprising, as recent C releases in an adjacent area were found to result mostly from new OC sources (Schuur et al., 2009). This could suggest that

even with enhanced vegetative production, increases in the relative abundances of genes involved in carbohydrate metabolism are nonetheless more attributable to soil temperature, at least in frigid locations where microbial catabolic functions are constrained by cold conditions.

Large differences in broad community features were observed between 15-25 cm and 45-55 cm communities in functional and taxonomic composition (Figure 5-1; Figure 5-2), average genome size (Table D-2), and estimated sequence-based community diversity (Figure D-1). A previous assessment based on short-term laboratory incubations of active and permafrost boundary layer communities, observed how permafrost soil community structure converged towards that of active layer communities (Mackelprang et al., 2011). The opposite was observed in our *in situ* field warming experiment despite reflecting a much larger time frame for adaptation to occur (Figure 5-2). The β-diversity distance between 45-55 cm and 15-25 cm communities were significantly greater for warmed 45-55 cm soils (paired t-test, $P < 0.001$ for abundance weighted Jaccard distances of KO terms; $P < 0.05$ for weighted UniFrac of 16S OTUs clustered at 97% similarity). This result could also be due to redox conditions that became more constraining at 45-55 cm as a result of warming (from enhanced electron acceptor consumption).

5.6. Conclusions and future work

Our evaluation revealed a much greater susceptibility of recently-thawed permaftost soil communities to further thaw and temperature increases, as well as responses that constrasted with warming-induced changes to pre-existing active layer communities. Metabolic shift in the recently-thawed permafrost community reflected a stimulation of methanogens, particularly those using

acetate for CH_4 production. These responses could be attributable to plant-microbe interactions arising from certain plant species that were also stimulated by warming, such as *Eriophorum vaginatum*. Functional responses also inferred a shift in redox conditions, which could be due to enhanced use of electron acceptors with declining redox potentials and can also help explain the large response in methanogenic taxa. These effects were not soley attributable to increased thaw depth or duration, but also to the independent effects of temperature. If accelerated warming results in more thorough depletion of electron acceptors than would occur under a more gradual thaw, it could result in a greater CH_4:CO_2 of emissions. Related to this, increasing CH_4 emissions observed in this area now outweigh the warming potential of CO_2 emissions and more than offset C uptake by local plants (Natali et al., 2015; Taylor et al., 2018). Meanwhile, northern-latitude areas have experienced a rate of warming over the past several decades that is much greater than the global average. If responses under gradual vs. rapid warming are dissimilar in potential CH_4 release, it could imply that inferences made from natural thaw gradients or geological records may not serve as adequate predictors of future CH_4 release from tundra soils if warming continues at a rapid pace in northern latitudes. As the rate of warming could be a primary determinant of the global warming potential of tundra soil C emissions, this topic should remain a focus of future research and modeling efforts. Shifts observed at the 15-25 cm layer were distinctly attributable to functions involved in carbohydrate metabolisms, occurred gradually over time, and were most attributable to the direct effect of warming. These results also suggest that tundra communities not undergoing a critical phase shift may respond more narrowly to warming.

While an increase in the abundance of methanogenic taxa and genes certainly implies that these taxa were active and generated CH_4 at some point, numerous environmental conditions and biological factors (i.e., plants, methanotrophs) play a role in methane emission vs. consumption. Thus, knowledge of when activity occurred is important for understanding how susceptible these soils are to CH_4 loss and to better predict future CH_4 emissions. Assessments of *in-situ* microbial activity, such as through the use of proteomics and transcriptomics of permafrost soils (Hultman et al., 2015; Coolen and Orsi 2015; Mackelprang et al., 2016), can offer such insights regarding community functioning. The emerging breadth of approaches making soil microbiota accessible for investigation should further serve to unravel the complex responses and interactions between ecological attributes in warming tundra ecosystems; thus, leading in overall greater understanding of how the they might respond and contribute to ongoing climate change. The gene and genome sequences reported here should facilitate proteomics and primer design for PCR assays that can be used to precisely monitor the abundance dynamics of responsive populations in tundra and elsewhere, and further corroborate these results.

5.7. Acknowledgements

This research was supported by the U.S. Department of Energy (award DE-SC0004601 to JZ, ZH, LW, YL, EAGS, JMT, and KTK) and the US National Science Foundation (award 1356288 to KTK).

5.8. References

Abbott BW, Jones JB, Schuur EAG, et al (2016) Biomass offsets little or none of permafrost carbon release from soils, streams, and wildfire: an expert assessment. Environmental Research Letters 11:034014. doi: 10.1088/1748-9326/11/3/034014

An S, Gardner W (2002) Dissimilatory nitrate reduction to ammonium (DNRA) as a nitrogen link, versus denitrification as a sink in a shallow estuary (Laguna Madre/Baffin Bay, Texas). Marine Ecology Progress Series 237:41–50. doi: 10.3354/meps237041

Bais HP, Weir TL, Perry LG, et al (2006) The role of root exudates in rhizosphere interactions with plants and other organisms. Annual Review of Plant Biology 57:233–266. doi: 10.1146/annurev.arplant.57.032905.105159

Barter RL, Yu B (2018) Superheat: An R Package for Creating Beautiful and Extendable Heatmaps for Visualizing Complex Data. Journal of Computational and Graphical Statistics 1–30. doi: 10.1080/10618600.2018.1473780

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57:289–300

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59–60. doi: 10.1038/nmeth.3176

Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: architecture and applications. BMC bioinformatics 10:421

Cantarel BL, Coutinho PM, Rancurel C, et al (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research 37:D233–D238. doi: 10.1093/nar/gkn663

Caporaso JG, Kuczynski J, Stombaugh J, et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335–336. doi: 10.1038/nmeth.f.303

Ciais P, Sabine C, Bala G et al. (2013) Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (eds Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM) pp Page, Cambridge, United Kingdom and New York, NY, USA.

Coolen MJL, Orsi WD (2015) The transcriptional response of microbial communities in thawing Alaskan permafrost soils. Frontiers in Microbiology 6:. doi: 10.3389/fmicb.2015.00197

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485. doi: 10.1186/1471-2105-11-485

DeSantis TZ, Hugenholtz P, Larsen N, et al (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology 72:5069–5072. doi: 10.1128/AEM.03006-05

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. doi: 10.1093/bioinformatics/btq461

Elberling B, Michelsen A, Schädel C, et al (2013) Long-term CO2 production following permafrost thaw. Nature Climate Change 3:890–894. doi: 10.1038/nclimate1955

Giblin A, Tobias C, Song B, et al (2013) The Importance of Dissimilatory Nitrate Reduction to Ammonium (DNRA) in the Nitrogen Cycle of Coastal Ecosystems. Oceanography 26:124–131. doi: 10.5670/oceanog.2013.54

GISTEMP Team, 2018: GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies. Dataset accessed 2018-07-11 at https://data.giss.nasa.gov/gistemp/.

Grogan P, Jonasson S (2003) Controls on annual nitrogen cycling in the understorey of a sub-arctic birch forest. Ecology 84:202-218

Hallam SJ (2004) Reverse Methanogenesis: Testing the Hypothesis with Environmental Genomics. Science 305:1457–1462. doi: 10.1126/science.1100025

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Reviews of Geophysics 48:. doi: 10.1029/2010RG000345

Heimann M, Reichstein M (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451:289–292. doi: 10.1038/nature06591

Hicks Pries CE, Schuur EAG, Crummer KG (2012) Holocene Carbon Stocks and Carbon Accumulation Rates Altered in Soils Undergoing Permafrost Thaw. Ecosystems 15:162–173. doi: 10.1007/s10021-011-9500-4

Hicks Pries CE, Schuur EAG, Crummer KG (2013) Thawing permafrost increases old soil and autotrophic respiration in tundra: Partitioning ecosystem respiration using δ 13 C and ∆ 14 C. Global Change Biology 19:649–661. doi: 10.1111/gcb.12058

Hinzman LD, Deal CJ, McGuire AD, et al (2013) Trajectory of the Arctic as an integrated system. Ecological Applications 23:1837–1868. doi: 10.1890/11-1498.1

Hultman J, Waldrop MP, Mackelprang R, et al (2015) Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature 521:208–212. doi: 10.1038/nature14238

Hyatt D, Chen G-L, LoCascio PF, et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. doi: 10.1186/1471-2105-11-119

IPCC (Intergov. Panel Clim. Change). 2013. Near-term climate change: projections and predictability. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, ed. TF Stocker, D Qin, GK Plattner, M Tignor, SK Allen, et al., pp. 953–1028. Cambridge, UK: Cambridge Univ. Press

Jain C, Rodriguez-R LM, Phillippy AM, et al (2017) High-throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. doi: 10.1101/225342

Johnston ER, Rodriguez-R LM, Luo C, et al (2016) Metagenomics Reveals Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem. Frontiers in Microbiology 7:. doi: 10.3389/fmicb.2016.00579

Jorgenson MT, Racine CH, Walters JC, Osterkamp TE (2001) Permafrost degradation and ecological changes associated with a warming climate in central Alaska. Climatic Change 48:551–579

Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. doi: 10.7717/peerj.1165

Kao-Kniffin J, Freyre DS, Balser TC (2010) Methane dynamics across wetland plant species. Aquatic Botany 93:107–113. doi: 10.1016/j.aquabot.2010.03.009

Keuper F, Bodegom PM, Dorrepaal E, et al (2012) A frozen feast: thawing permafrost increases plant-available nitrogen in subarctic peatlands. Global Change Biology 18:1998–2007. doi: 10.1111/j.1365-2486.2012.02663.x

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217. doi: 10.1093/bioinformatics/bts611

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. doi: 10.1038/nmeth.1923

Lawrence DM, Slater AG (2005) A projection of severe near-surface permafrost degradation during the 21st century. Geophysical Research Letters 32:. doi: 10.1029/2005GL025080

Lawrence DM, Slater AG, Swenson SC (2012) Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4. Journal of Climate 25:2207–2225. doi: 10.1175/JCLI-D-11-00334.1

Lee H, Schuur EAG, Inglett KS, et al (2012) The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. Global Change Biology 18:515–527. doi: 10.1111/j.1365-2486.2011.02519.x

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Lipson DA, Haggerty JM, Srinivas A, et al (2013) Metagenomic Insights into Anaerobic Metabolism along an Arctic Peat Soil Profile. PLoS ONE 8:e64659. doi: 10.1371/journal.pone.0064659

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:. doi: 10.1186/s13059-014-0550-8

Mackelprang R, Saleska SR, Jacobsen CS, et al (2016) Permafrost Meta-Omics and Climate Change. Annual Review of Earth and Planetary Sciences 44:439–462. doi: 10.1146/annurev-earth-060614-105126

Mackelprang R, Waldrop MP, DeAngelis KM, et al (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480:368–371. doi: 10.1038/nature10576

Mu C, Zhang T, Wu Q, et al (2015) Editorial: Organic carbon pools in permafrost regions on the Qinghai–Xizang (Tibetan) Plateau. The Cryosphere 9:479–486. doi: 10.5194/tc-9-479-2015

McCalley CK, Woodcroft BJ, Hodgkins SB, et al (2014) Methane dynamics regulated by microbial community response to permafrost thaw. Nature 514:478–481. doi: 10.1038/nature13798

Natali SM, Schuur EAG, Rubin RL (2012) Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost: Increased plant productivity in Alaskan tundra. Journal of Ecology 100:488–498. doi: 10.1111/j.1365-2745.2011.01925.x

Natali SM, Schuur EAG, Trucco C, et al (2011) Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra: WARMING OF ALASKAN TUNDRA. Global Change Biology 17:1394–1407. doi: 10.1111/j.1365-2486.2010.02303.x

Natali SM, Schuur EAG, Mauritz M, et al (2015) Permafrost thaw and soil moisture driving CO 2 and CH 4 release from upland tundra. Journal of Geophysical Research: Biogeosciences 120:525–537. doi: 10.1002/2014JG002872

Nayfach S, Pollard KS (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biology 16:. doi: 10.1186/s13059-015-0611-7

Osterkamp TE, Jorgenson MT, Schuur EAG, et al (2009) Physical and ecological changes associated with warming permafrost and thermokarst in Interior Alaska. Permafrost and Periglacial Processes 20:235–256. doi: 10.1002/ppp.656

Parks DH, Imelfort M, Skennerton CT, et al (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25:1043–1055. doi: 10.1101/gr.186072.114

Parks DH, Rinke C, Chuvochina M, et al (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology 2:1533–1542. doi: 10.1038/s41564-017-0012-7

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. doi: 10.1093/bioinformatics/bts174

Penton CR, Yang C, Wu L, et al (2016) NifH-Harboring Bacterial Community Composition across an Alaskan Permafrost Thaw Gradient. Frontiers in Microbiology 7:. doi: 10.3389/fmicb.2016.01894

Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2018) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5:e9490. doi: 10.1371/journal.pone.0009490

Quast C, Pruesse E, Yilmaz P, et al (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41:D590–D596. doi: 10.1093/nar/gks1219

Rodriguez-R LM, Gunturu S, Harvey WT, et al (2018a) The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Research 46:W282–W288. doi: 10.1093/nar/gky467

Rodriguez-R LM, Gunturu S, Tiedje JM, et al (2018b) Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems 3:e00039-18. doi: 10.1128/mSystems.00039-18

Rodriguez-R LM, Konstantinidis KT (2014) Estimating coverage in metagenomic data sets and why it matters. Isme Journal 8:2349–2351. doi: 10.1038/ismej.2014.76

Romanovsky VE, Smith SL, Christiansen HH (2010) Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007-2009: a synthesis. Permafrost and Periglacial Processes 21:106–116. doi: 10.1002/ppp.68

Sistla SA, Schimel JP (2013) Seasonal patterns of microbial extracellular enzyme activities in an arctic tundra soil: Identifying direct and indirect effects of long-term summer warming. Soil Biology and Biochemistry 66:119–129. doi: 10.1016/j.soilbio.2013.07.003

Schuur EAG, Abbott B (2011) Climate change: High risk of permafrost thaw. Nature 480:32–33. doi: 10.1038/480032a

Schuur EAG, Abbott BW, Bowden WB, et al (2013) Expert assessment of vulnerability of permafrost carbon to climate change. Climatic Change 119:359–374. doi: 10.1007/s10584-013-0730-7

Schuur EAG, Bockheim J, Canadell JG, et al (2008) Vulnerability of Permafrost Carbon to Climate Change: Implications for the Global Carbon Cycle. BioScience 58:701–714. doi: 10.1641/B580807

Schuur EAG, Vogel JG, Crummer KG, et al (2009) The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. Nature 459:556–559. doi: 10.1038/nature08031

Ström L, Mastepanov M, Christensen TR (2005) Species-specific Effects of Vascular Plants on Carbon Turnover and Methane Emissions from Wetlands. Biogeochemistry 75:65–82. doi: 10.1007/s10533-004-6124-1

Su X, Pan W, Song B, et al (2014) Parallel-META 2.0: Enhanced Metagenomic Data Analysis with Functional Annotation, High Performance Computing and Advanced Visualization. PLoS ONE 9:e89323. doi: 10.1371/journal.pone.0089323

Tarnocai C, Canadell JG, Schuur EAG, et al (2009) Soil organic carbon pools in the northern circumpolar permafrost region: SOIL ORGANIC CARBON

POOLS. Global Biogeochemical Cycles 23:n/a-n/a. doi: 10.1029/2008GB003327

Taylor MA, Celis G, Ledman JD, et al (2018) Methane efflux measured by eddy covariance in Alaskan upland tundra undergoing permafrost degradation. Journal of Geophysical Research: Biogeosciences. doi: 10.1029/2018JG004444

Thauer RK, Shima S (2008) Methane as Fuel for Anaerobic Microorganisms. Annals of the New York Academy of Sciences 1125:158–170. doi: 10.1196/annals.1419.000

The UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic Acids Research 43:D204–D212. doi: 10.1093/nar/gku989

Wang Y, Inamori R, Kong H, et al (2008) Influence of plant species and wastewater strength on constructed wetland methane emissions and associated microbial populations. Ecological Engineering 32:22–29. doi: 10.1016/j.ecoleng.2007.08.003

Wu Y-W, Simmons BA, Singer SW (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. doi: 10.1093/bioinformatics/btv638

Xue K, M. Yuan M, J. Shi Z, et al (2016) Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. Nature Climate Change 6:595–600. doi: 10.1038/nclimate2940

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. doi: 10.1093/bioinformatics/btt593

Zhang X, Johnston ER, Li L, et al (2017) Experimental warming reveals positive feedbacks to climate change in the Eurasian Steppe. The ISME Journal 11:885–895. doi: 10.1038/ismej.2016.180

APPENDIX A: SUPPLEMENTAL MATERIAL FOR CHAPTER 2

A.1. Supplementary methods and materials

A.1.1. Soil chemical and physical determination

For each unique soil collection, soil moisture was determined by measuring the weight loss of ~10g field moist soil dried at 105 ºC for 48 hours (calculated as: 1 - dry soil weight/wet soil weight). The average moisture content of two independently dried subsamples was used for analysis. Soil pH was determined using an automated LabFit AS-3000 pH Analyzer and soil extractable P, K, Ca, Mg, Mn, and Zn were extracted using the Mehlich-1 method and measured using an inductively coupled plasma spectrograph at the University of Georgia Agricultural and Environmental Services Laboratories (Athens, GA, USA). Soil extractable P using this method is interpreted as the available fraction of P. Measurement of NH_4-N and NO_3-N was also performed at the same facility. For this, samples were first extracted with 0.1 N KCl, following the phenate method for NH_4^+ and the cadmium reduction method for NO_3^-, and measurements were determined colorimetrically.

Soil texture analysis was performed using a slightly modified version of the Bouyoucos Hydrometer method outlined in Methods of Soil Analysis part 1 (Gee et al., 2002). In brief, 40 g subsample of soil was combined with 100 ml of 0.082 M hexametaphosphate and 250 ml of deionized water and allowed to soak overnight. Then, soil was blended in an electric mixer for 5 minutes. The mixture was then transferred to a 1 L beaker and the volume was adjusted to 1 L with

deionized water. A plunger was used to suspend the soil particles in the solution and measurements were taken with a hydrometer.

Total N (TN) and total C (TC) were measured using a Shimadzu TOC-L CSH/CSN analyzer. For total N, 24 ml samples are placed on an auto sampler; each sample is introduced to the combustion tube (furnace temperature 720 °C), where TN in the sample decomposes to become nitrogen monoxide. The carrier gas, which contains the nitrogen monoxide, is cooled and dehumidified by an electric dehumidifier. Then it enters a chemiluminescesnce gas analyzer, where the nitrogen monoxide is detected. The detection signal from the chemiluminescence gas analyzer generates a peak and the TN concentration in the sample was measured. For TC, carrier gas flows at 150 mL/min into the combustion tube, which has been filled with an oxidation catalyst and heated to 680 °C. The TC of a sample is burned in the combustion tube to produce carbon dioxide. The carrier gas, containing the carbon dioxide and other combustion products, flows from the combustion tube to a dehumidifier (electronic dehumidifier), where it is cooled and dehydrated. It then passes through a halogen scrubber before it reaches the cell of a non-dispersive infrared NDIR gas analyzer, where the carbon dioxide is detected. The analog detection signal of the NDIR forms a peak, and the area of this peak is measured by an internal data processor. The peak area is proportional to the TC concentration of the sample; thus, when a TC standard solution has been analyzed to create a calibration curve, i.e., the equation expressing the relationship between TC concentration and peak area, the TC concentration in the sample can be calculated. The Kjeldhal Analysis method was used for determination of total soil P (TP). For each sample, 0.5 grams of air-dried and sieved soil was ground and digested with H_2SO_4 acid and Kjeltab (1.5 g K_2SO_4 + 0.15 g $CuSO_4$) in a block digester at

390 °C. The resulting orthophosphate was analyzed to quantify TP using a Flow Injection Analyzer based on the Lachat Method No 13-115-01-1-B (Lachat Instruments, Loveland, CO, USA).

Microbial biomass carbon (MBC) and microbial biomass nitrogen (MBN) were measured using the chloroform fumigation-extraction method (Vance et al., 1987). The non-fumigated sample was combined with 0.5 M K_2SO_4 added at a 7:1 solution to dry soil ratio, followed by shaking for 1 hr in a horizontal shaker at 200 rpm, and filtration through Whatman #1 filter paper. The non-fumigated sample represents the amount of dissolved C or N available in the soil. The second subsample was fumigated with ethanol-free chloroform for 3 days. After venting the residual chloroform, the sample was extracted using K_2SO_4 as described above. All filtrates were stored in the freezer until analysis with a Shimadzu Total Organic Carbon Analyzer (Model TOC-LCSH) and Total Nitrogen Analyzer (Model TNM-L). K_2SO_4 blanks were also analyzed in triplicate. The difference in C or N between the fumigated and non-fumigated samples gives the chloroform-labile C or N pool (EC or EN). MBC was calculated using the equation MBC = EC/k where k is estimated as 0.45 (Beck et al., 1997), and MBN was calculated using the equation MBC = EC/k where k is estimated as 0.54 (Brookes et al. 1985). For determination of soil microbial biomass phosphorus (MBP), 10 g of soil was added to a 50 ml centrifuge tube with 36 ml of Bray solution (containing 0.03 M NH_4F and 0.025 M HCl) and 4 ml of DI water. Mixtures were placed in a horizontal shaker at 200 rpm for 1 hour, and then filtered using Whatman #42 filter paper. As with MBC and MBN determination, subsamples were fumigated with ethanol-free chloroform for 3 days. The difference in Bray-extractable P from fumigated and non-fumigated samples was taken to represent soil MBP. For MBC, MBP, and MBN determination, the

average of two independently processed and measured subsamples from each soil collection was used for analysis.

A.1.2. Soil enzyme assays

Enzyme assays were performed on field, control incubation, and P-amended incubation soil samples by measuring the activities on target substrates labeled with 4-methylumbelliferone (MUB). See Figure A-2 for a diagram of incubation soils used for enzyme activity analysis. The substrates in our study included 4-MUB-α-D-glucopyranoside for α-glucosidase (AG), 4-MUB-β-D-glucopyranoside for β-glucosidase (BG), 4-MUB-β-D-cellobioside for β-D-cellubiosidase (CB), 4-MUB-N-acetyl-β-D-glucosaminide for N-acetyl β-glucosaminidase (NAG), 4-MUB-β-D-xylopyranoside for β-xylosidase (XYL), 4-MUB-phosphate for phosphomonoesterase (PHOS), and bis-4-MUB-phophate for phosphodiesterase (DIPHOS). For each soil sample, soil equivalent to 2.75 g of dry weight was measured. Sodium acetate buffer (50 mM) was adjusted to match the pH of each soil sample. The soil sample was subsequently mixed with 91 mL of the pH-adjusted sodium acetate buffer. The soil slurry was divided into three technical replicates. Each well in the 96-well plate was filled by 800 µl of the soil slurry. For standard plates, each row of the wells was further mixed with 200 µl of MUB solutions of the concentrations of 2.5, 5, 10, 25, 50, 100 µM respectively. For enzyme reaction plate, each row of the wells was mixed with 200 µl of each MUB labeled substrate (200 µM). The incubation at 26 ºC was performed on both standard and enzyme reaction plates for 3 hours. Subsequently, the plates were taken out of the incubator and centrifuged for 8 min at 1500 rpm. From each well, 250 µl of the supernatant was transferred to a clean reading plate. The reaction was terminated by adding 5 µl of 0.5 M NaOH.

Measurements for reading plates were conducted on Tecan Infinite M200 microplate reader for the absorbance at 365 nm. The enzyme activities were quantified as nmol/gram-dry-soil/hour based on the regression lines of those standard curves. Relative phosphomonoesterase activity was comparable between soils. Thus, enzyme activities were evaluated as a ratio between SOC-degrading enzyme activities (namely, those of α-glucosidase, β-glucosidase, β-D-cellubiosidase, N-acetyl β-glucosaminidase, and β-xylosidase) and phosphodiesterase (DIPHOS) activity (e.g., β-glucosidase:DIPHOS), an approach used in previous studies to compare enzyme activities between soils (Sinsabaugh et al., 2009; Ramirez et al., 2012; Turner and Wright 2014).

A.2.1. Supplemental references

Beck T;, Joergensen R.G., Kandeler E., et al., 1997. An inter-laboratory comparison of ten different ways of measuring soil microbial biomass C. Soil Biology and Biochemistry 29, 1023–1032.

Brookes P.C., Landman A., Pruden G., Jenkinson D.S., 1985. Chloroform fumigation and the release of soil nitrogen: A rapid direct extraction method to measure microbial biomass nitrogen in soil. Soil Biology and Biochemistry 17, 837–842.

Gee G.W., Or D., 2002. Particle-size analysis. In: Dane JH and Topp GC, editors, Methods of soil analysis. Part 4. Physical methods. SSSA Book Ser. 5. SSSA, Madison, WI. p. 255–293.

Ramirez K.S., Craine J.M., Fierer N., 2012. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. Global Change Biology 18, 1918–1927.

Sinsabaugh R.L., Hill B.H., Shah J.J.F., 2009. Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. Nature 462, 795.

Turner B.L., Wright S.J., 2014. The response of microbial biomass and hydrolytic enzymes to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. Biogeochemistry 117, 115–130.

Vance E.D., Brookes P.C., Jenkinson D.S., 1987. Microbial biomass measurements in forest soils: The use of the chloroform fumigation-incubation method in strongly acid soils. Soil Biology and Biochemistry 19, 697–702.

A.3. Supplemental figures

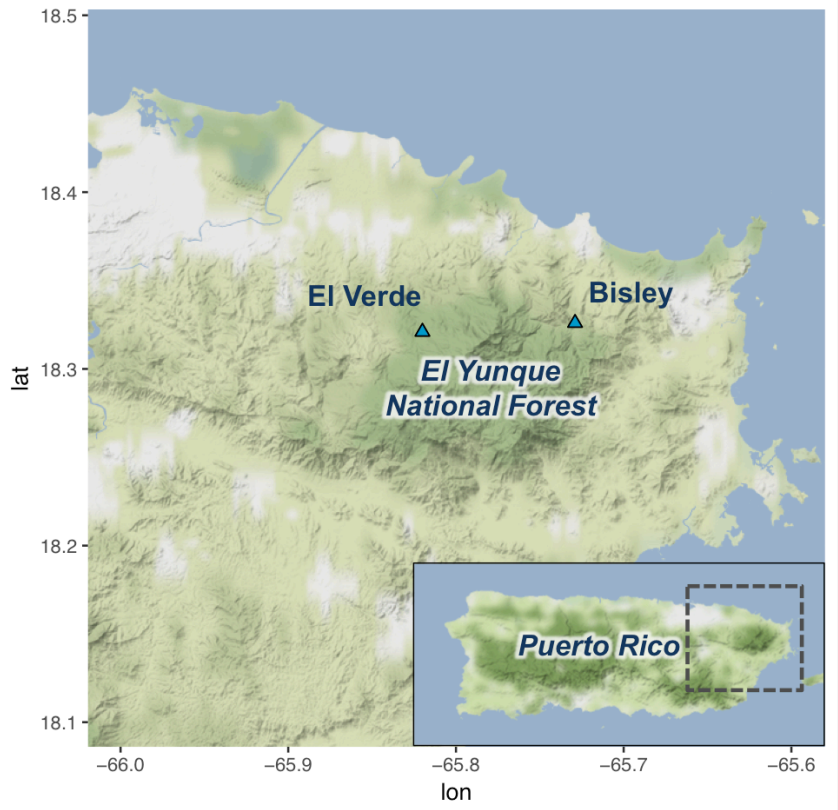


Figure A-1: Map of the field sites located in the El Yunque National Forest (Puerto Rico, USA).

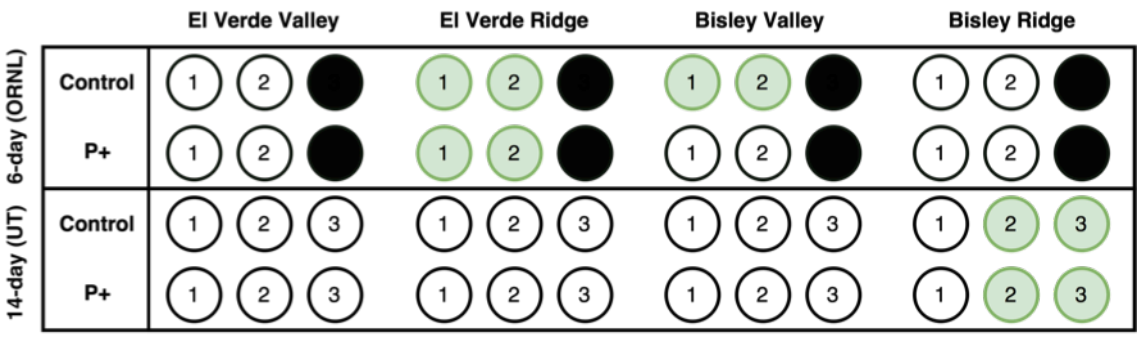


Figure A-2: Diagram of soil incubations performed for ~6 days and ~14 days using Oak Ridge National Laboratory (ORNL) and University of Tennessee (UT) MicroOxymax Respirometer systems (Columbus Instruments, Columbus, OH), respectively. Numbers in circles denote unique samplings from each of the four sites. Open white circles (not black) represent incubations that were performed for each of the two setups; black circles represent incubations not performed. Those colored in green were selected for metagenomics, metatranscriptomics, and enzyme assay analyses. High quality RNA from Bisley ridge sampling 1 and all El Verde soil samples soils could not be recovered.

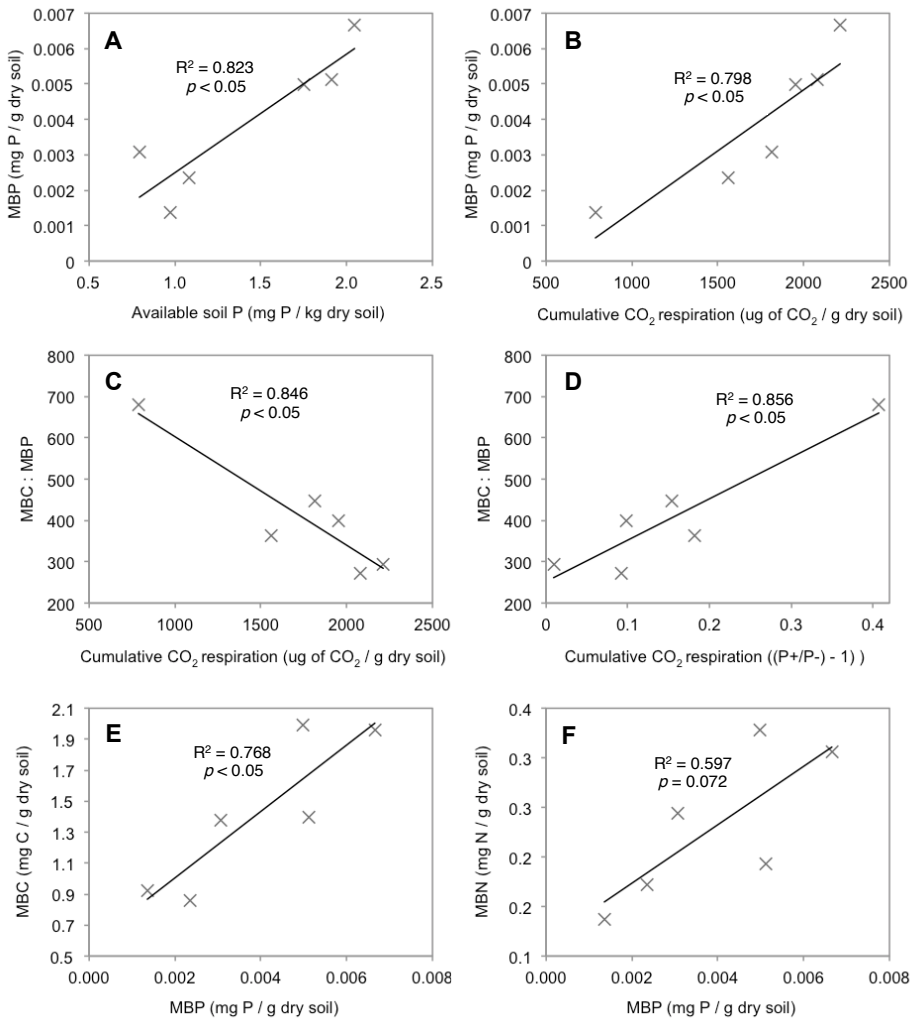


Figure A-3: Correlations between respiration activity and measured soil indices for ridge soils. Graphs show (A) available soil phosphorus vs. microbial biomass phosphorus (MBP), (B) cumulative CO_2 respiration after 14 days vs. MBP, (C) cumulative CO_2 respiration after 14 days vs. the ratio of microbial biomass carbon (MBC) to MBP, (D) the difference in CO_2 respiration after 14 days between control and P amended incubations vs. the ratio of MBC to MBP, (E) MBP vs. MBC, (F) MBP vs. microbial biomass nitrogen. Statistical significance of Pearson correlation was determined using two-tailed Student's t-distribution.

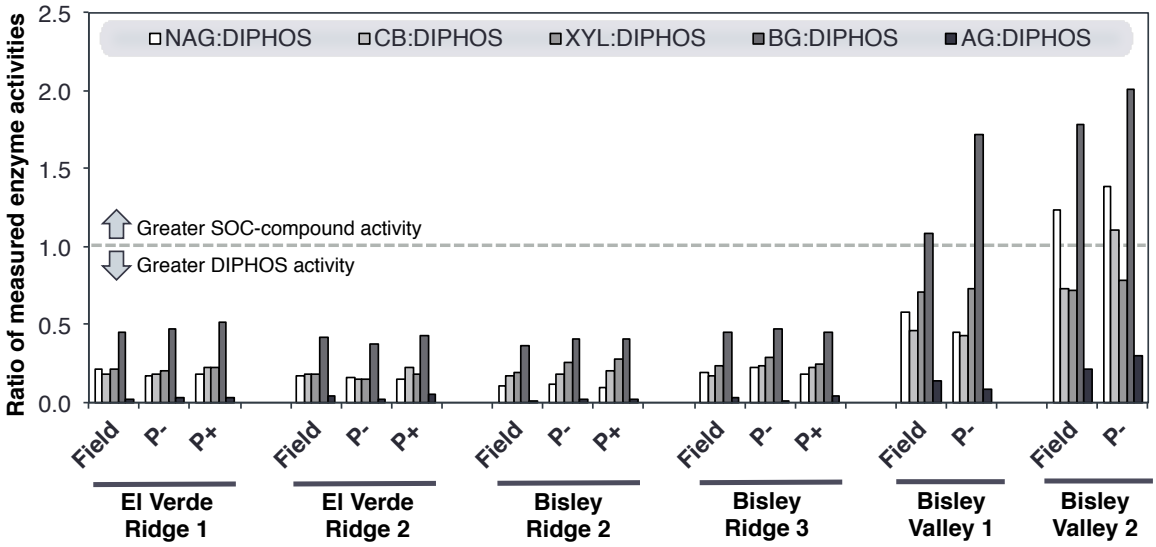


Figure A-4: Relative enzymatic activity for soils from field (Field), control (P-), and incubations amended phosphorus (P+). Values are expressed as a ratio between various soil organic carbon-specific enzyme activities relative to the activity of diphosphoesterase. Labels are AG:α-glucosidase, BG:β-glucosidase, CB:β-D-cellubiosidase NAG:N-acetyl-β-glucosaminidase, XYL:β-xylosidase, DIPHOS:Phospho-diesterase.

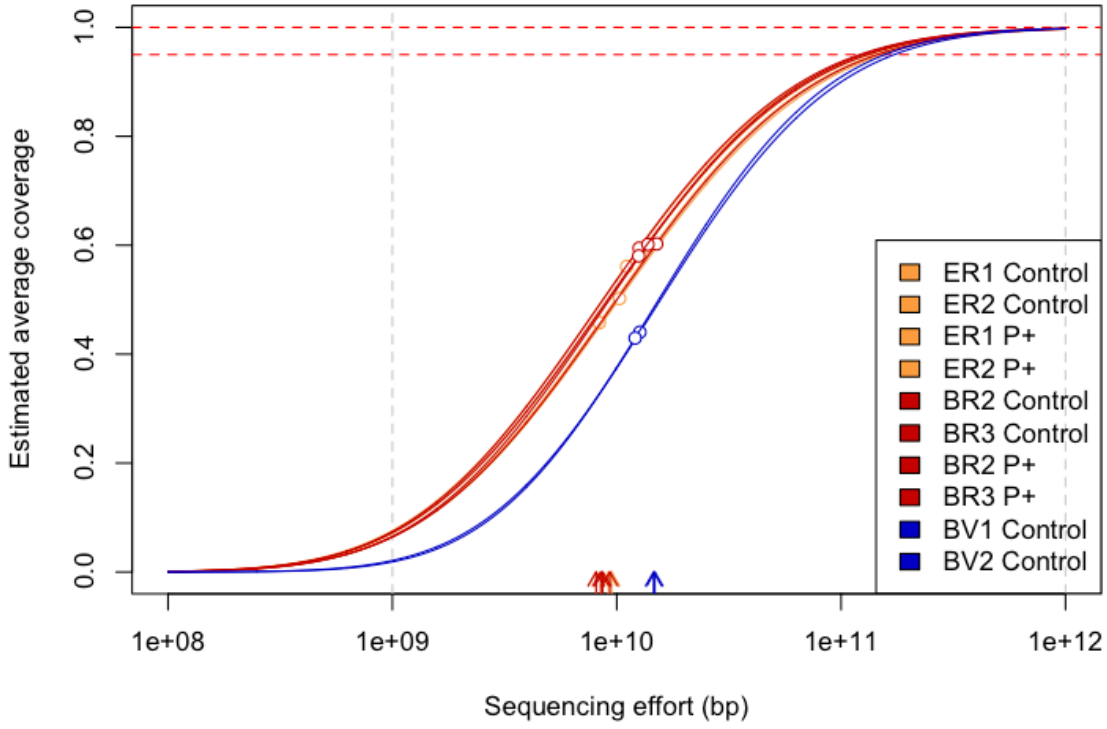


Figure A-5: Curves representing soil microbial community complexity as determined by Nonpareil. Nonpareil is a statistical tool that uses read redundancy to estimate metagenome dataset complexity and the amount of sequencing effort needed to achieve a desired level of coverage (Rodriguez-R *et al.,* 2018). Circles on curves represent the estimated average coverage at the sequencing depth/effort applied; projected line to the right of the circle represents the expected coverage for higher sequencing efforts. Dashed lines represent 95% and 99% coverage; curves positioned on the right represent more sequence-diverse metagenomes than curves positioned on the left.

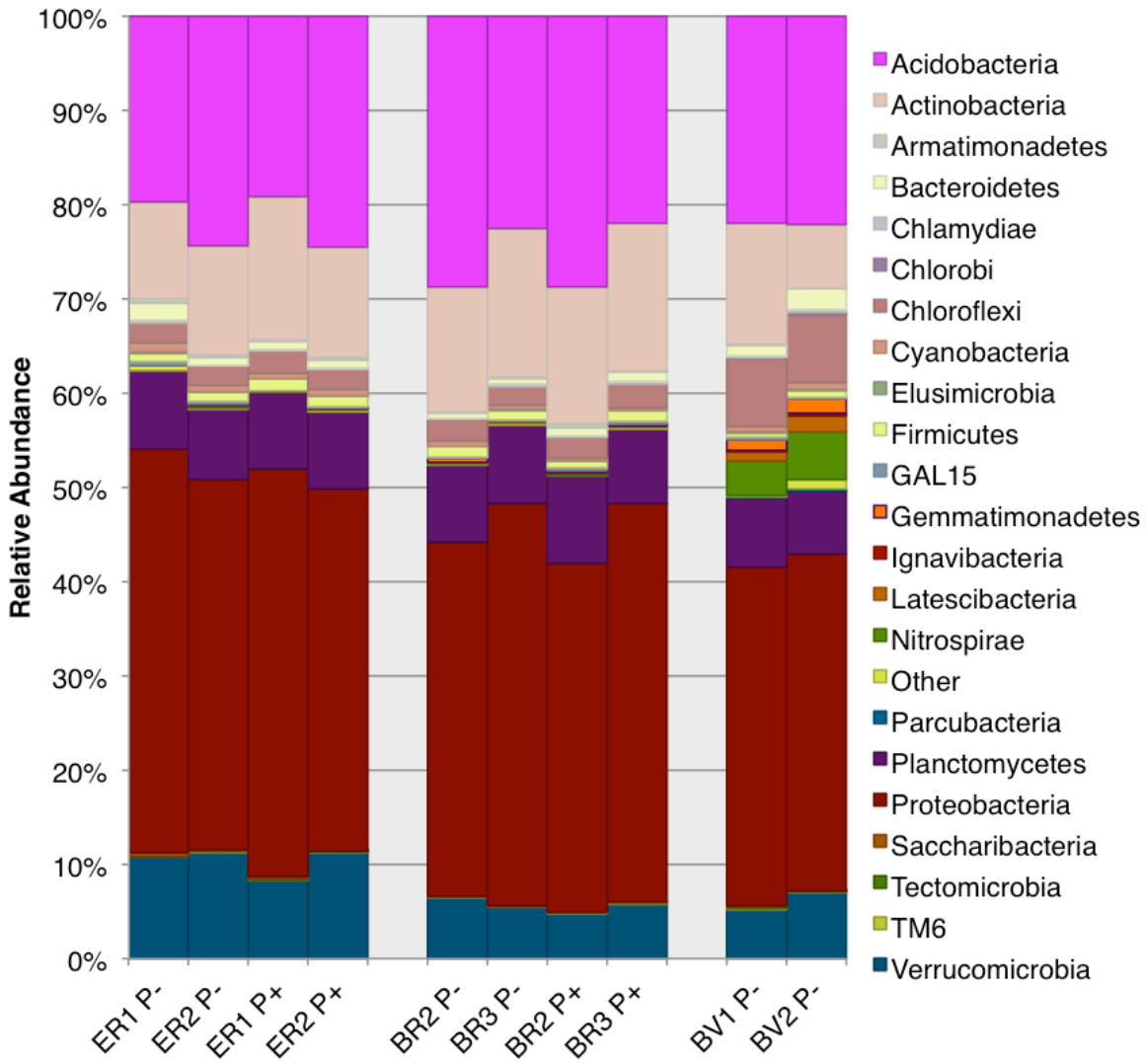


Figure A-6: Taxonomic composition (phylum-level) of soil communities, using 16S rRNA gene fragments recovered from each of the ten metagenomes used in our study.

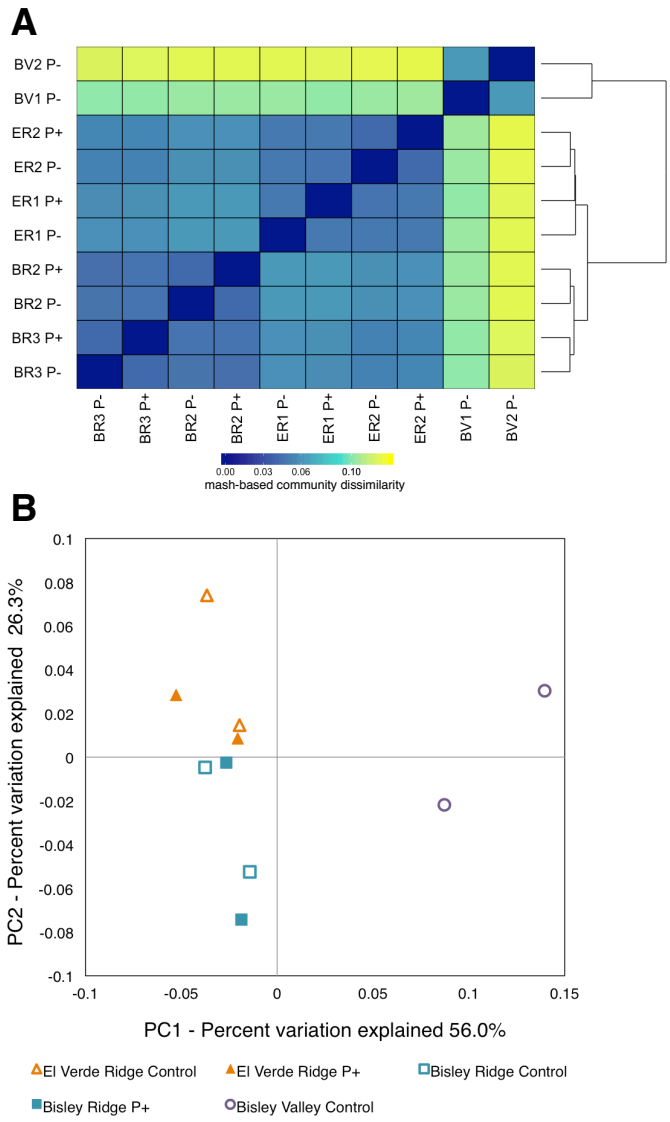


Figure A-7: (A) Heatmap showing the distances calculated with Mash (using kmer composition) between sample metagenomes. Heatmap was produced with superheat R package. Samples are ordered by hierarchical clustering, and a dendrogram is included to illustrate the arrangement of clusters produced by hierarchical clustering. **(B) PCoA plot illustrating the arrangement of samples based on taxonomic composition in the 2D plane spanned by their first two principal components**. Underlying data are Bray-Curtis distances calculated from phylum-level taxonomic composition.

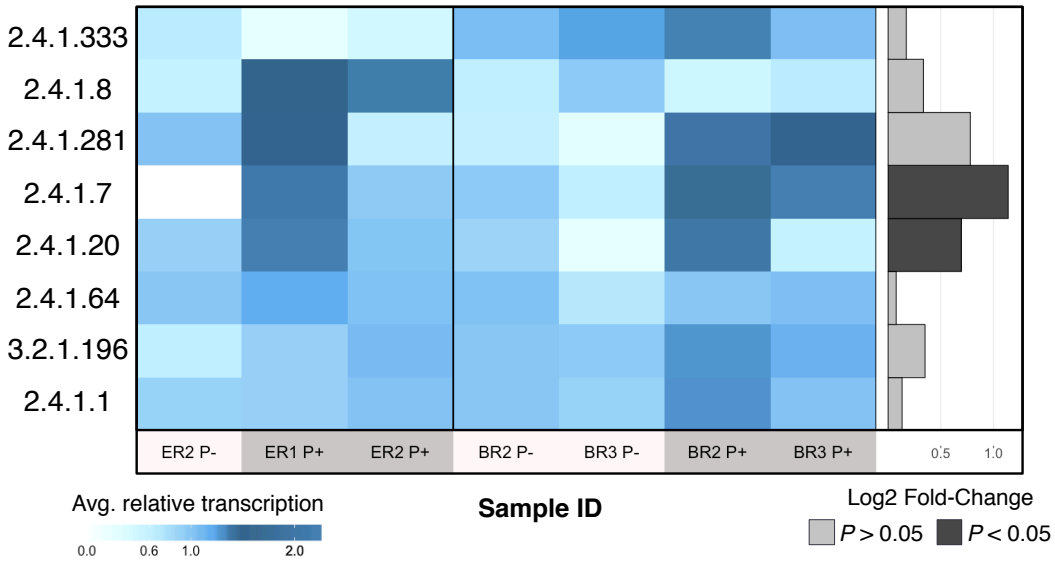


Figure A-8: Heatmap showing the relative transcription of phosphorylase genes and *glgX* for both control (P-) and corresponding P-amended incubations (P+) from El Verde ridge (ER) and Bisley ridge (BR) metatranscriptomes. Phosphorylase genes (summarized as enzyme commission IDs) incuded all entries represented by the SwissProt database involved in the phosphorolysis of carbohydrate substrates with an average relative abundance ≥ 1E-5. To better emphasize differences between datasets, values were normalized by the average relative transcript abundance of all datasets (by rows). Sub-plots on the right of each heatmap represent the log2 fold-change calculated for each gene obtained from edgeR, and are colored by statitistical significant differences between P+ and P- soil metatranscriptomes.

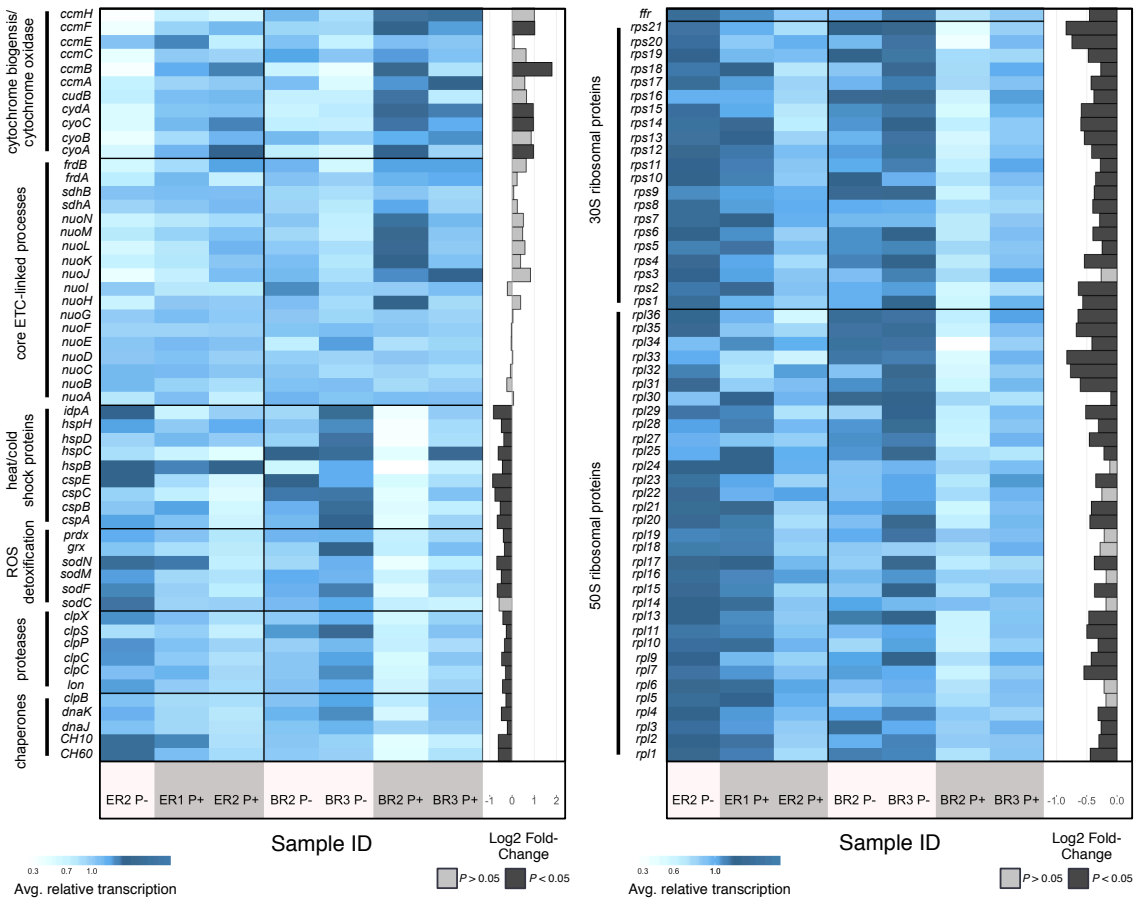


Figure A-9: Heatmap showing the relative transcription of genes related to electron transport chain (ETC) processes, cellular stress and maintenance, and ribosome biogenesis (bacterial ribosomal proteins) and recycling (*frr*) for both control (P-) and corresponding P-amended incubations (P+) from El Verde ridge (ER) and Bisley ridge (BR) metatranscriptomes. To better emphasize differences between datasets, values were normalized by the average relative transcript abundance of all datasets (by rows). Sub-plots on the right of each heatmap represent the log2 fold-change calculated for each gene obtained from edgeR, and are colored by statitistical significant differences between P+ and P- soil metatranscriptomes.

APPENDIX B: SUPPLEMENTAL MATERIAL FOR CHAPTER 3

B.1. Supplementary methods and materials

B.1.1. Shotgun metagenomic sequence pre-processing and annotation

Paired-end shotgun metagenomic sequences were merged using PEAR (Zhang *et al.*, 2014). Successfully merged reads underwent quality trimming using the SolexaQA package (Cox *et al.*, 2010); reads were trimmed where Phred quality scores dropped below 17 (≥98% accuracy). Nonmerge-able reads were trimmed separately. Reads ≥80-bp following trimming were retained. FragGeneScan was used for the protein prediction (Rho *et al*., 2010), adopting the Illumina 1% error model. The resulting amino acid (a.a.) sequences were searched against Swiss-Prot (UP consortium, 2014), using blastp (Camacho *et al.*, 2009) (blast+ version 2.2.29, options: –word_size 3, outfmt 6). Matches with a bit score >75, alignment length ≥25 a.a., and a.a. identity ≥40% were deemed acceptable for further analysis. A count matrix (with sample metagenomes as columns and gene annotations or metabolic process categories as rows) was generated to summarize the occurrence of each Swiss-Prot entry in each sample. Corresponding Gene Ontology Annotations of functions and processes for each Swiss-Prot entry was obtained from the uniprot_sprot.dat file provided on http://www.uniprot.org/downloads (downloaded on July 2014). A count matrix of summarized gene ontology pathways was processed with the DESeq2 package (Love *et al.*, 2014) to identify differentially-abundant pathways between the treatments and to generate log2 transformations of gene/process abundance

ratios. The raw count data underwent a variance-stabilizing transformation, which is used for logarithmically distributed count data with low mean values that tend to have high variance. This transformation results in new values that have a relatively constant variance along the range of mean values and confers a reduced false positive rate for less abundant genes (Anders & Huber, 2010). *P*-values of pathways for SOM catabolism and S and N cycling were transformed to account for false discovery rate from multiple testing using Benjamini–Hochberg correction (adjusted *P*-values; Benjamini and Hochberg, 1995).

B.1.2. Taxonomic composition analysis

Using both merged and non-merged reads ≥80bp after trimming, the relative abundance of various prokaryotic phyla was determined. To achieve this, candidate 16S rRNA sequences were first identified by searching reads from entire metagenomes against the May 2013 release of Greengenes 16S ribosomal database (DeSantis *et al.*, 2006) pre-clustered at 88% identity, using blastn (blast+ version 2.2.29, options: -word_size 16, -outfmt 6, -task blastn, -dust no). Matches with a bit-score ≥45 were used in a subsequent filtering process. Various scripts in Qiime were used for the majority of the remaining workflow (Caporaso *et al.*, 2010). Candidate 16S rRNA underwent closed-reference OTU (Operational Taxonomic Unit)-picking against Greengenes database pre-clustered at 99% (options: -m uclust_ref –s 0.97) (Edgar, 2010). Taxonomic annotations for each representative sequence provided by Greengenes were used for each resulting OTU. The phyla abundances for each sample were then summarized using this information.

To estimate the abundance of fungi relative to prokaryotes, these reads were also aligned to the Silva database (Quast *et al.*, 2013), using similar

procedures. In particular, 18S rRNA genes were used as references for fungi while 16S rRNA genes for archaea and bacteria. The percentage of reads belonging to each of the three groups was calculated as their relative abundance. Bacteria were found to make up ~94% of reads, and thus we further calculated bacterial taxonomic richness by sampling equal number of sequences.

B.2. Supplementary references

Anders S, Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biol* **11:** 106.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K *et al.* (2009). BLAST plus: architecture and applications. *BMC Bioinformatics* **10:** 421.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Cox MP, Peterson DA, Biggs PJ. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11:** 485.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.

Rho M, Tang H, Ye Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

UniProt (UP) consortium. (2014). Activities at the universal protein resource (UniProt). *Nucleic Acids Res* **42:** D191–D198.

Zhang J, Kobert K, Flouri T, Stamatakis A. (2014). PEAR: a fast and accurate Illumina paired-end read merge R. *Bioinformatics* **30:** 614–620.

B.3. Supplemental tables and figures

Watering	Warming	Sample ID	Number of raw sequences	Merged Reads used in blastp	Unique Nonmerged reads used in blastp	Total Reads used in Blastp	Annotations Made
0	0	ZXM-2	6884555	6595856	124994	6720850	974720
0	0	ZXM-3	7819079	6888224	518126	7406350	1148176
0	0	ZXM-4	9197718	6848079	1313986	8162065	1169287
0	0	ZXM-5	10170980	5698469	1967701	7666170	1096956
0	0	ZXM-7	14151791	10502887	723359	11226246	1632378
0	1	ZXM-8	12581286	7014906	2393488	9408394	1357400
0	1	ZXM-9	11380278	5587307	2934746	8522053	1182923
0	1	ZXM-10	13011134	9499393	806617	10306010	1522940
0	1	ZXM-11	12353161	9569367	529014	10098381	1475441
0	1	ZXM-12	13543460	9311805	1231322	10543127	1557768
0	1	ZXM-13	30159789	17040454	6098582	23139036	3091273
1	0	ZXM-14	13486585	9295828	986281	10282109	1575095
1	0	ZXM-16	11865752	9391361	495893	9887254	1414182
1	0	ZXM-17	13550713	10890587	365306	11255893	1611091
1	0	ZXM-18	11252495	8517501	839109	9356610	1364387
1	0	ZXM-19	11152412	8935248	391014	9326262	1255115
1	0	ZXM-20	9999200	7826047	535233	8361280	1321991
1	1	ZXM-21	8657718	7674047	224799	7898846	1233259
1	1	ZXM-22	8483799	7500108	244559	7744667	1239844
1	1	ZXM-23	8539432	7404922	333403	7738325	1210705
1	1	ZXM-24	9112796	7943643	361350	8304993	1290767
1	1	ZXM-25	12553267	8710502	1977398	10687900	1595095
1	1	ZXM-26	14116212	9027687	3090373	12118060	1609336
Mean (standard deviation)			11914070 (4532103)	8594532 (2319260)	1238550 (1372003)	9833082 (3236203)	1431745 (407992)

Table B-1: Information pertaining to the merge-ability, sequence quality, and number of annotation made for shotgun-metagenome datasets representing each soil sample. The 0 and 1 for watering/warming represent without and with the treatment, respectively.

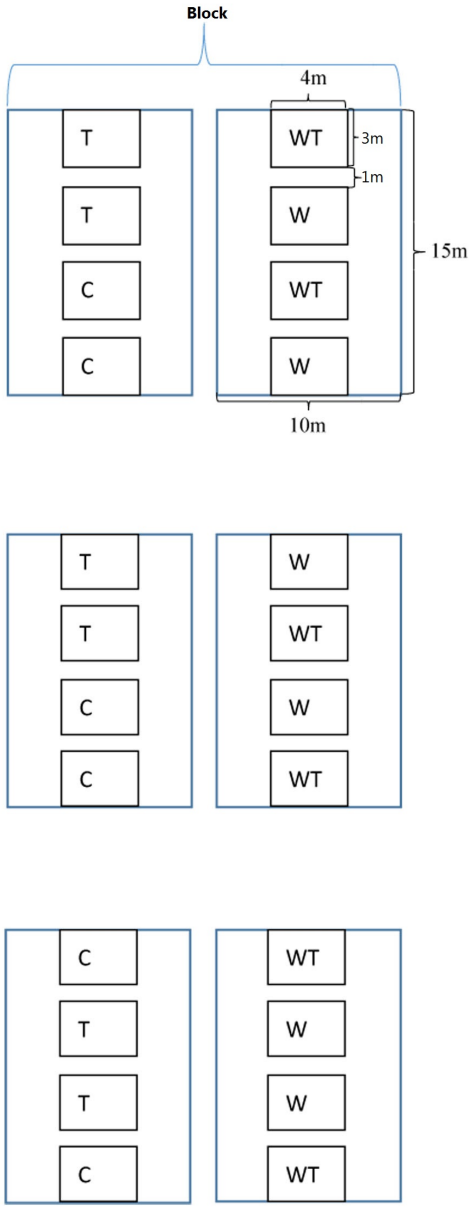


Figure B-1: The spatial distribution of plots in the three blocks. C, T, W and WT means control, warming, watering and warming plus watering, respectively.

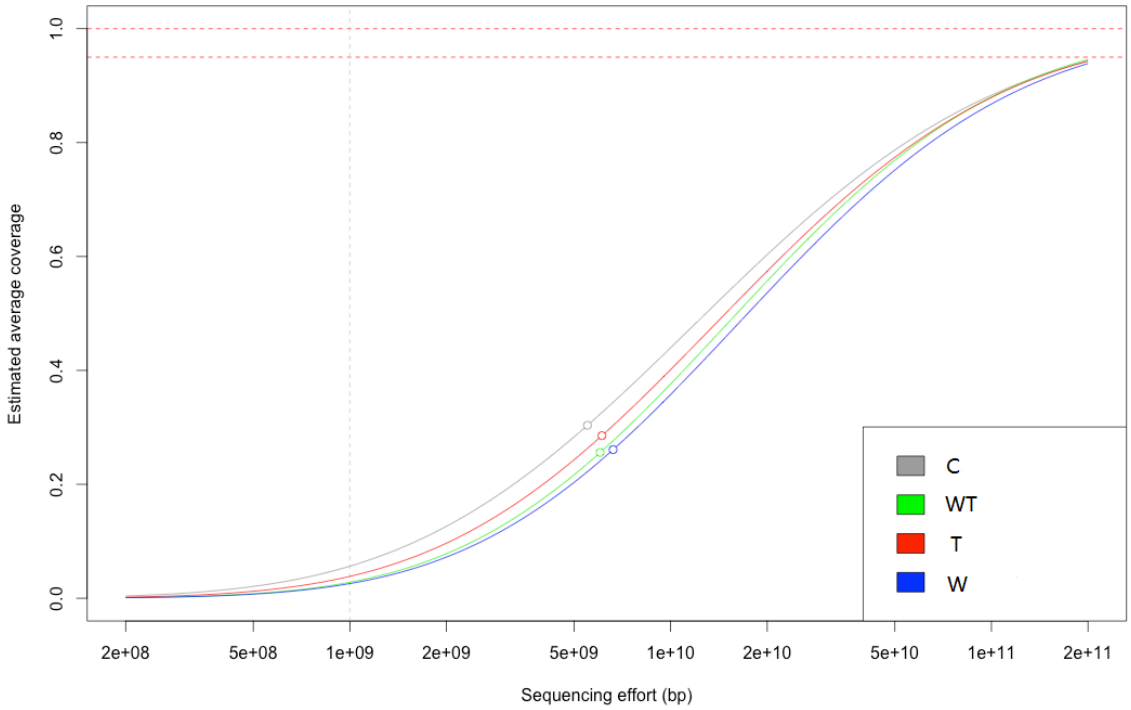


Figure B-2: Curves representing soil microbial community complexity estimations as determined by Nonpareil. Nonpareil is a statistical tool that uses read redundancy to estimate dataset complexity and the amount of sequencing effort needed to achieve a desired level of coverage. Circles on curves represent the coverage of the actual sequencing richness from the dataset supplied by the user in relation to the entire curve. Curves positioned on the right represent more sequence diverse metagenomes than curves positioned on the left.

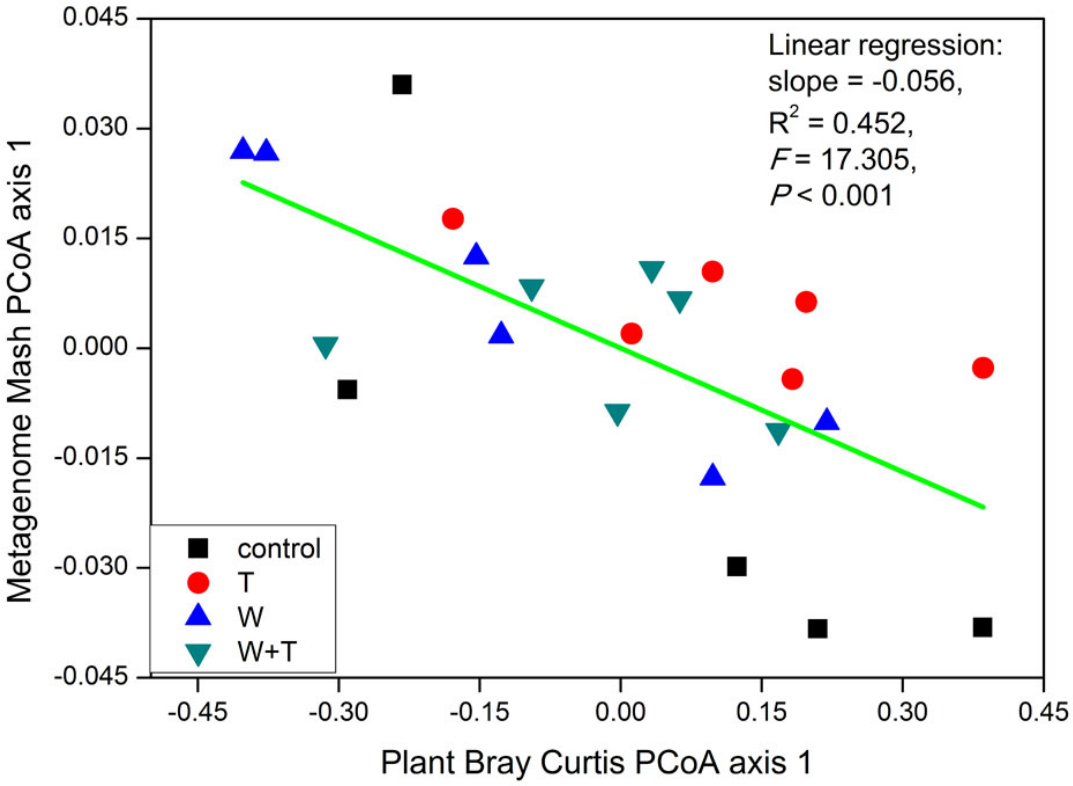


Figure B-3: Correlation between changes in microbial community and plant community compositions as an effect of the treatments. The graph represents the primary axes resulting from PCoA of Microbial community Mash-based distances (y-axis) and Bray-Curtis distances of the corresponding plant communities (x-axis) for the same experimental plots. The proportion of variation explained by Plant Bray Curtis PCoA axis 1 is 37.2% and by metagenomics Mash PCoA axis 1 is 7.5%.

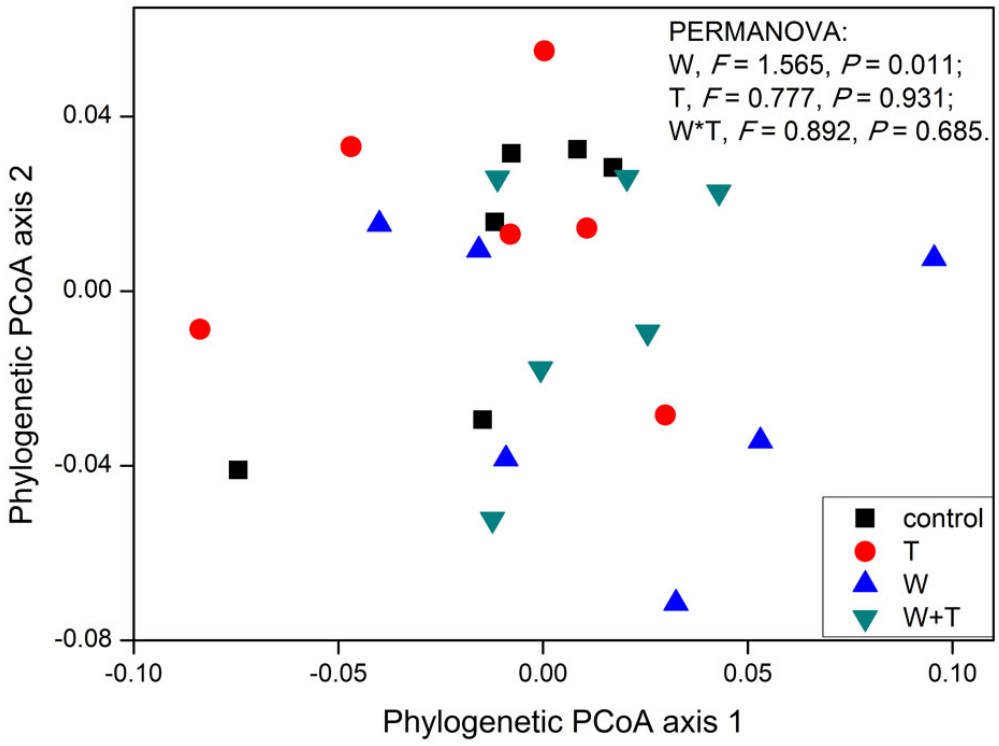


Figure B-4: Effect of experimental treatments on the phylogenetic composition of soil microbial communities. The graph represents the results of the PCoA analysis of the weighted UniFrac distances among the communities sampled based on 16S rRNA gene amplicon-based OTUs. The proportion of variation explained by PCoA axes 1 and 2 is 14.11% and 9.82%, respectively.

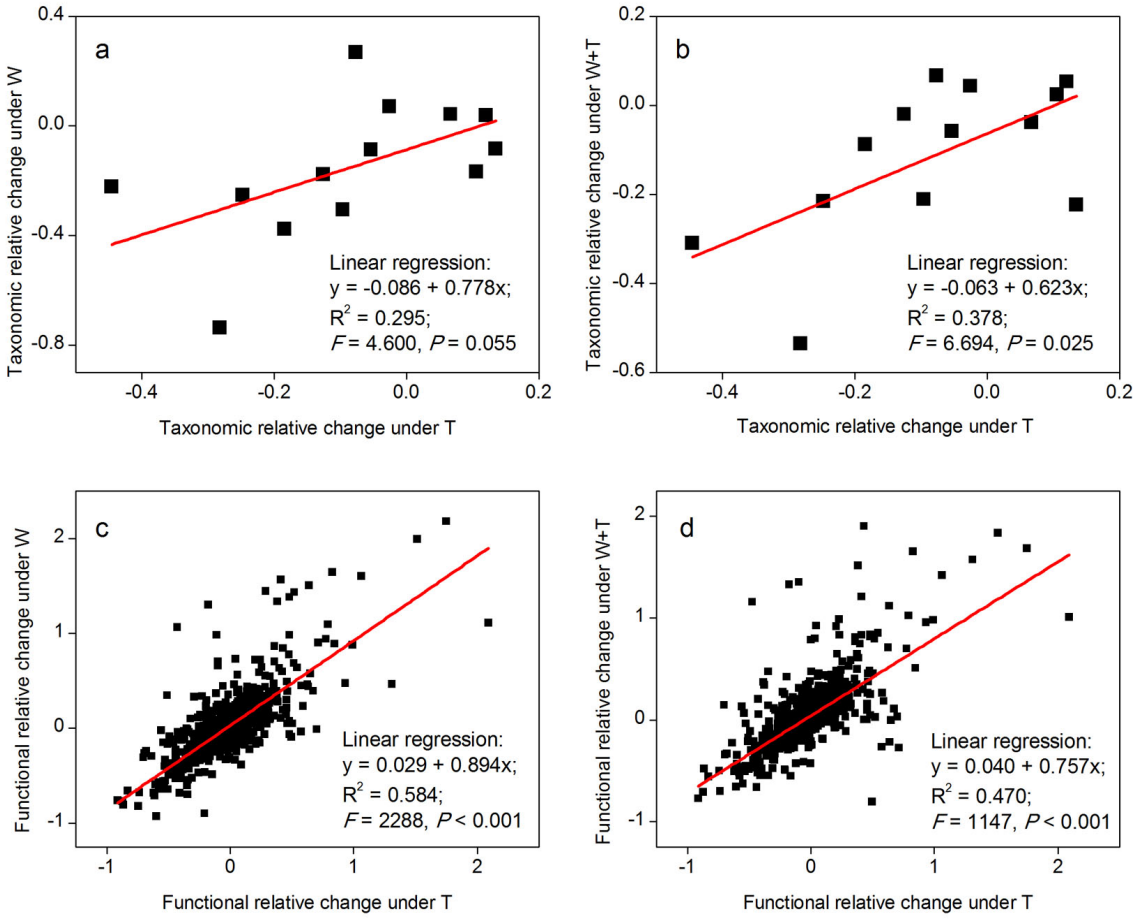


Figure B-5: The similar effects of different experimental treatments on the taxonomic (a, b) and functional (c, d) structure of microbial communities.

APPENDIX C: SUPPLEMENTAL MATERIAL FOR CHAPTER 4

C.1. Supplemental tables and figures

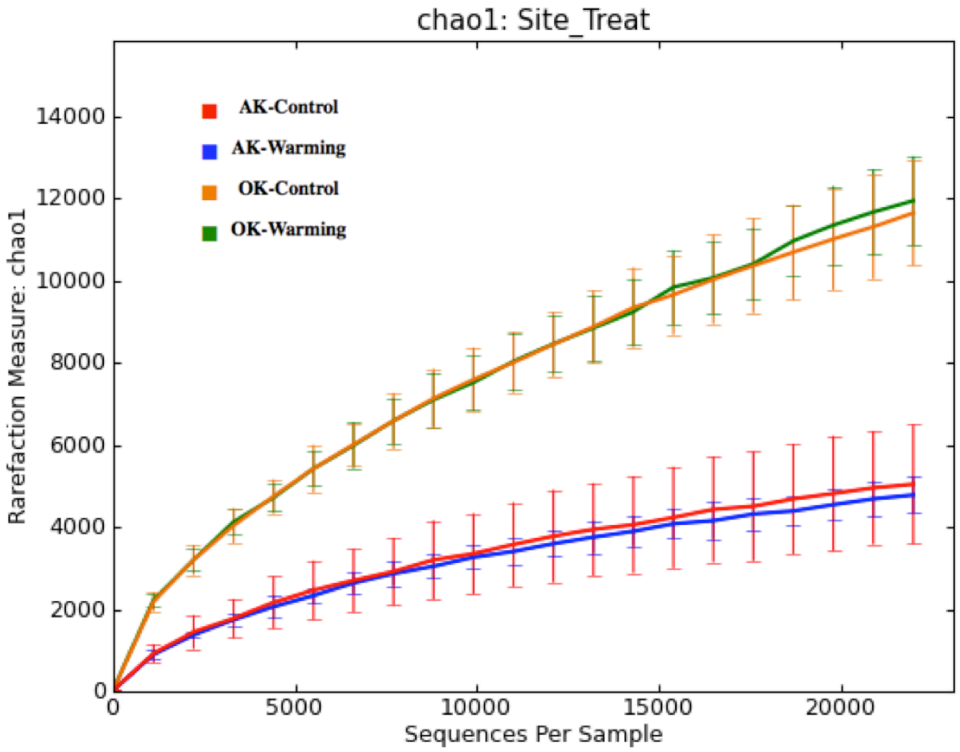


Figure C-1: OTU rarefaction curve. Displays the number of OTUs obtained by subsampling at different sequencing depths; an estimate of microbial community diversity. Estimates are obtained by random sampling of a supplied OTU table of 16S rDNA gene amplicons from region V4 clustered at 97% similarity, with removal of singleton OTUs. Chao1 metric was used for alpha diversity measurement.

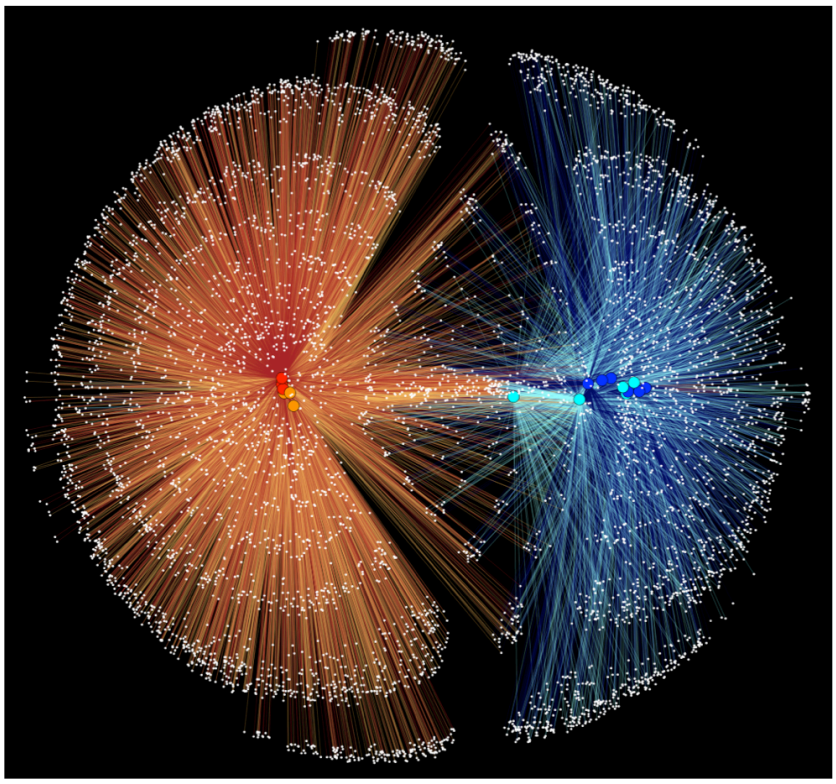


Figure C-2: OTU sharing network representing OTUs (PCR amplicons from 16S rRNA gene clustered at 97% similarity). Samples are clustered (positioned) according to the presence and abundance of their shared OTUs (using make_otu_network.py, a QIIME script; Caporaso et al. 2010b). Bold red and orange dots represent Oklahoma soil samples, warming and control plots, respectively; Bold dark blue and light blue dots represent Alaska soil samples, warming and control, respectively. White dots represent OTUs, and a line connecting these dots to a bolded sample dot indicates that the OTU is present in that sample. After removal of singleton OTUs from the dataset (OTUs appearing only once), 8,290 OTUs are represented at the OK site and 6,292 OTUs are represented at the AK site, with only 766 OTUs shared between the two sites.

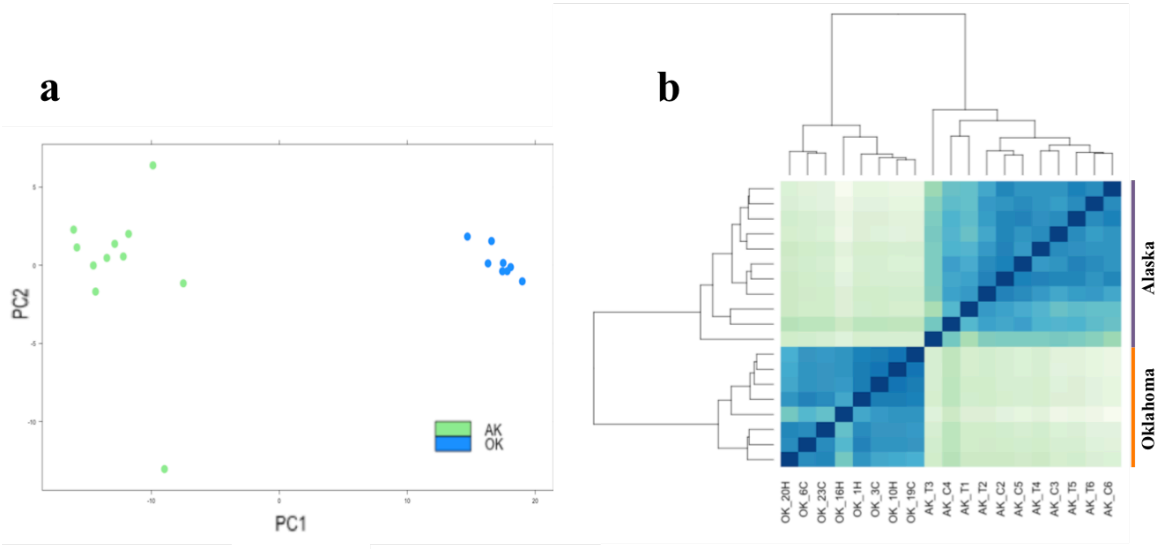


Figure C-3: (**a**) **PCA plot.** The 19 samples shown in the 2D plane spanned by their first two principal components. Underlying data are gene count matrix (short read annotation to Swiss-Prot) which have undergone variance stabilizing transformation in DESeq2 package. (**b**) **Sample-to-sample distances**. Heatmap showing the Euclidean distances between samples as calculated from the variance stabilizing transformation in DESeq2.

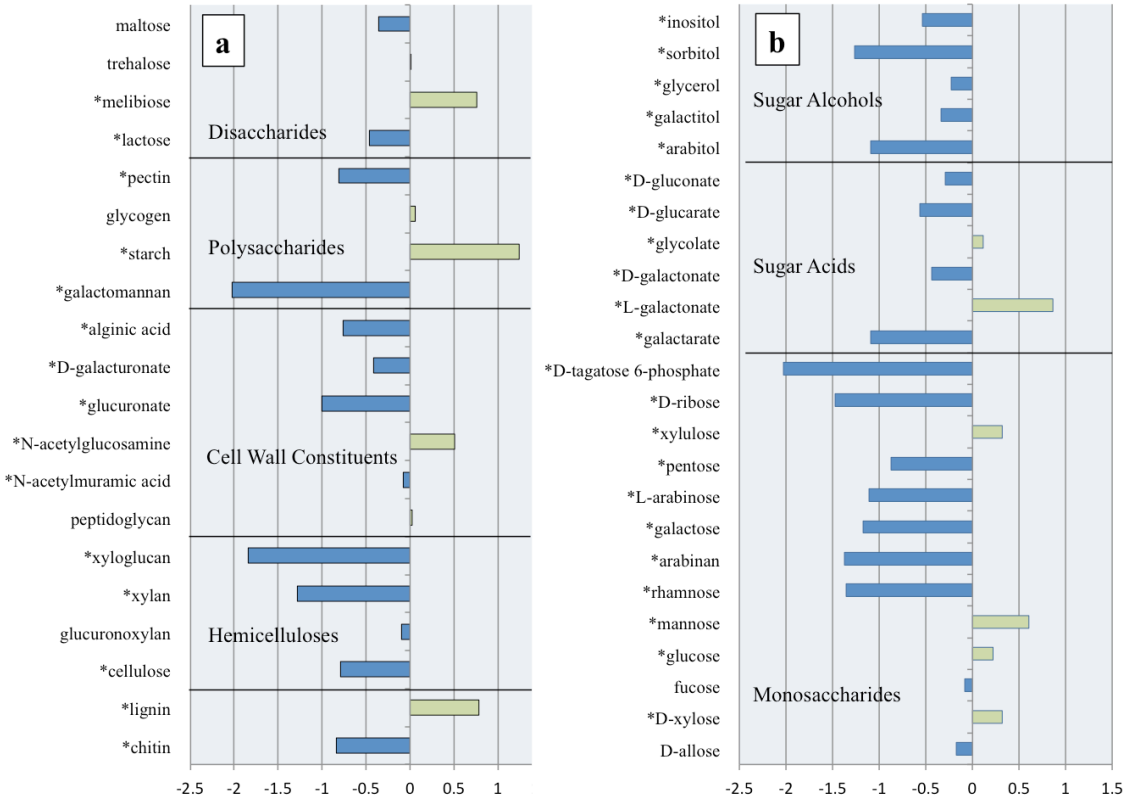


Figure C-4: Log2-fold differences in gene abundance between OK and AK soil samples of select SOM catabolic pathways. Dark grey bars represent pathways that were more abundant in AK metagenomes and light grey bars represent pathways that were more abundant in OK metagenomes. Pathways include (a) disaccharides, polysaccharides, cell wall constituents, hemicelluloses, cellulose, lignin, and chitin, a well as (b) sugar alcohols, sugar acids, and monosaccharides. SOM constituent catabolic pathways preceded with an asterisk denote those that were significantly different between sites (p-value < 0.001, DESeq2).

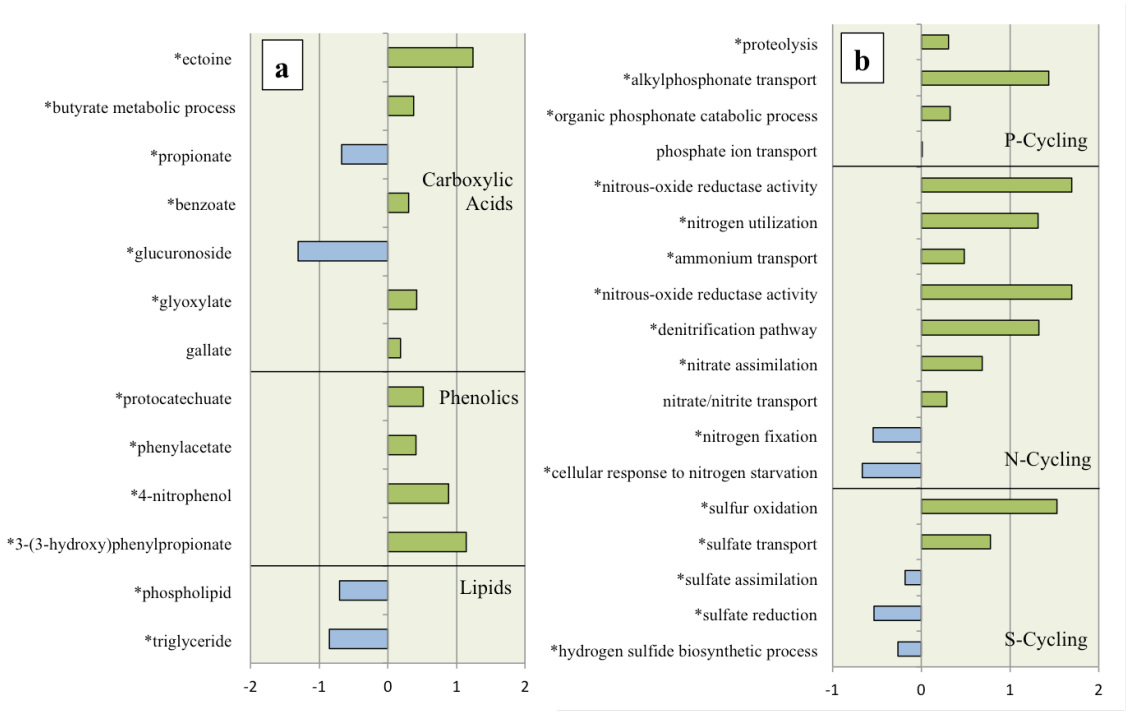


Figure C-5: Log2-fold differences in gene abundance between OK and AK soil samples for (a) catabolic pathways of lipids, phenolic compounds, and carboxylic acids, as well as (b) select S, N, and P-cycle pathways. Blue bars represent pathways that were more abundant in AK metagenomes and green bars represent pathways that were more abundant in OK metagenomes. Pathways names preceded with an asterisk denote those that were significantly different between sites (p-value < 0.001, DESeq2).

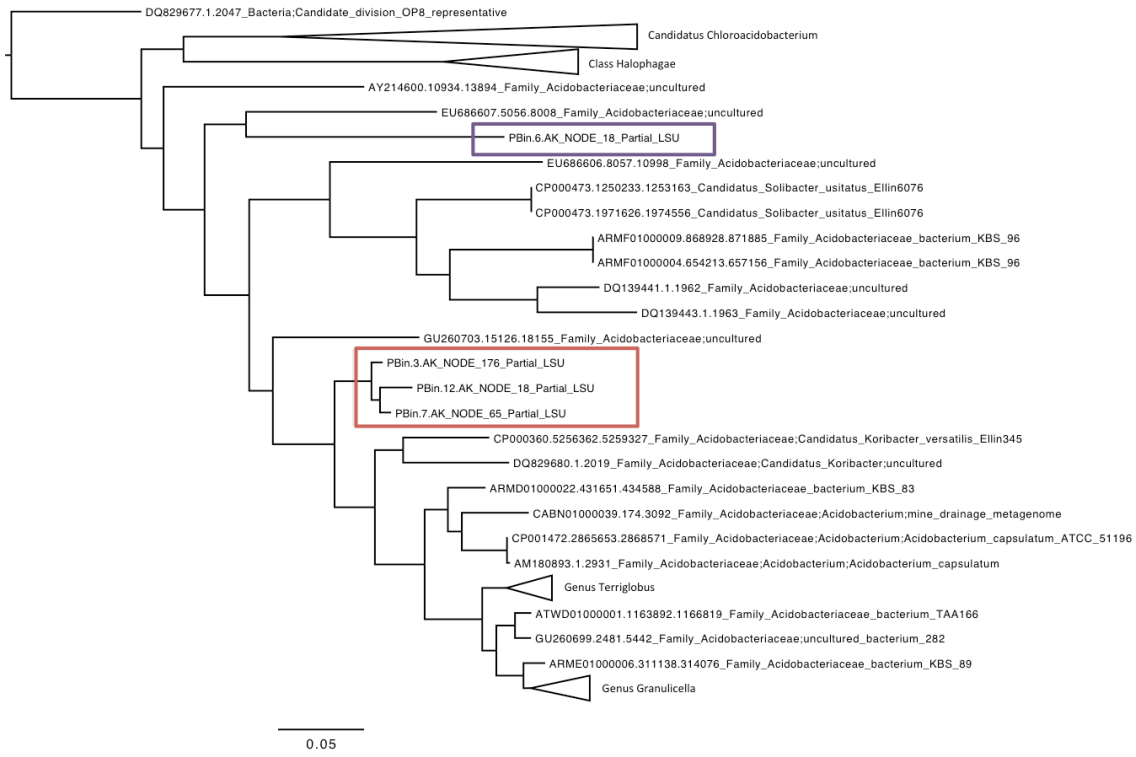
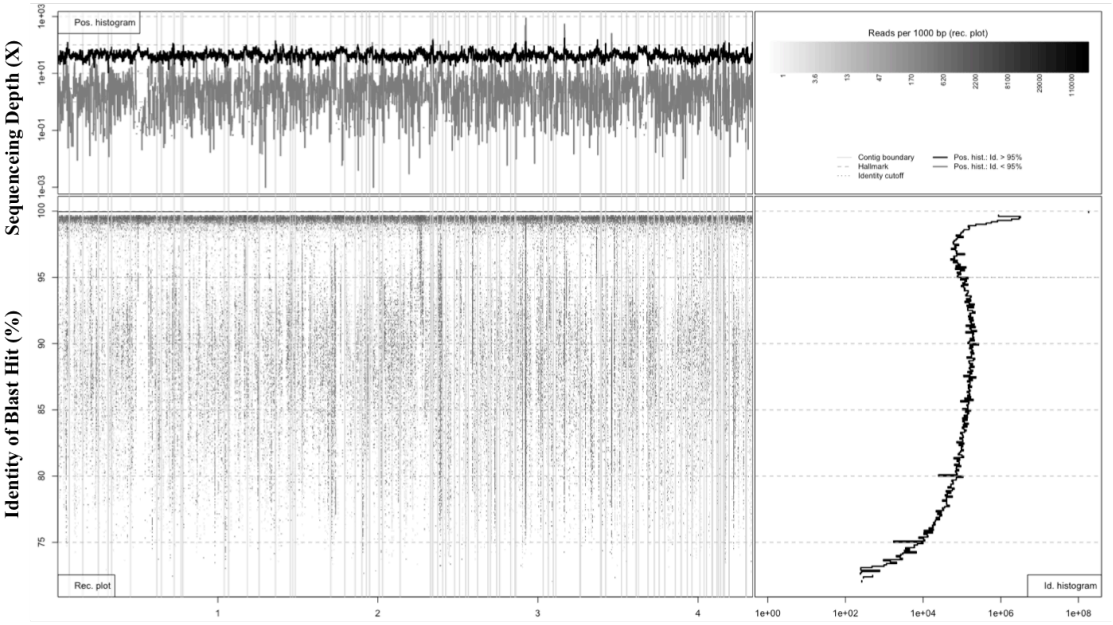
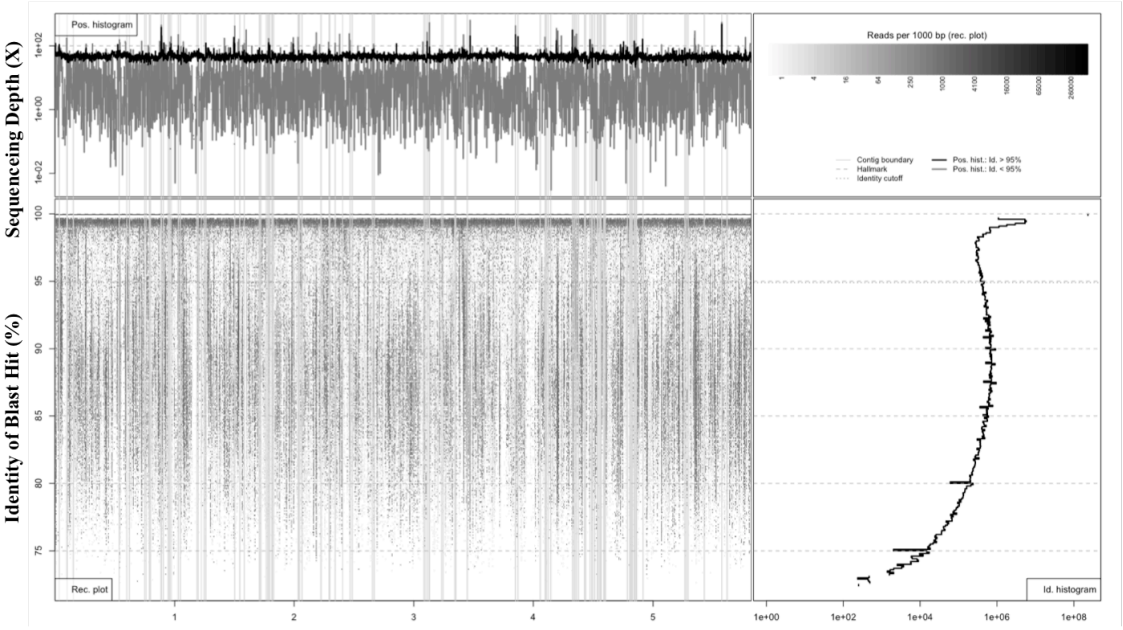


Figure C-6: Phylogenetic tree displaying relatedness of bin assemblies 03, 06, 07, and 11/12 to other known *Acidobacteria* representatives. Underlying data are aligned LSU sequences from bins and acidobacterial LSU sequences from SILVA LSU database (Quast et al. 2013) using a representative from *Ca. OP8* as an out-group. For many sequences, the SILVA name was modified for simplicity/aesthetics and unique accession numbers are provided in all instances that link many nodes to respective database sequences. Monophyletic groups representing bins are outlined in red (bins 03, 07, 11/12*)* and purple (bin 06).

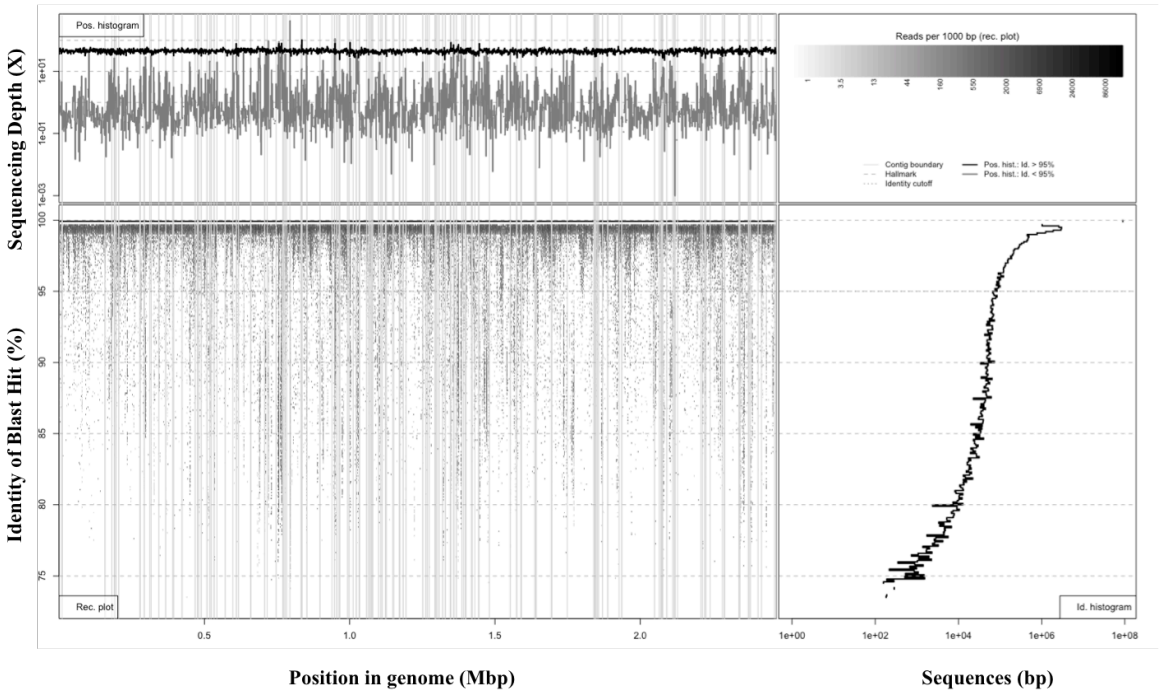
Fragment Recruitment of Bin 01 vs. AK-C2 sample metagenomic dataset



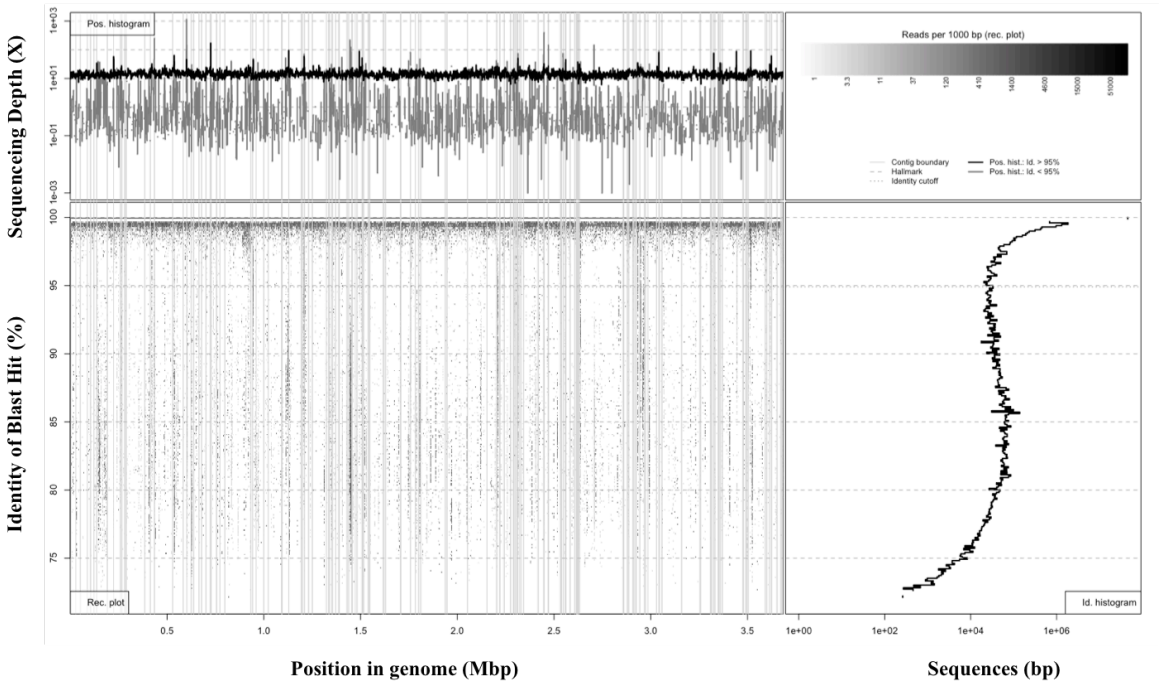
Fragment Recruitment of Bin 03 vs. AK-T5 sample metagenomic dataset



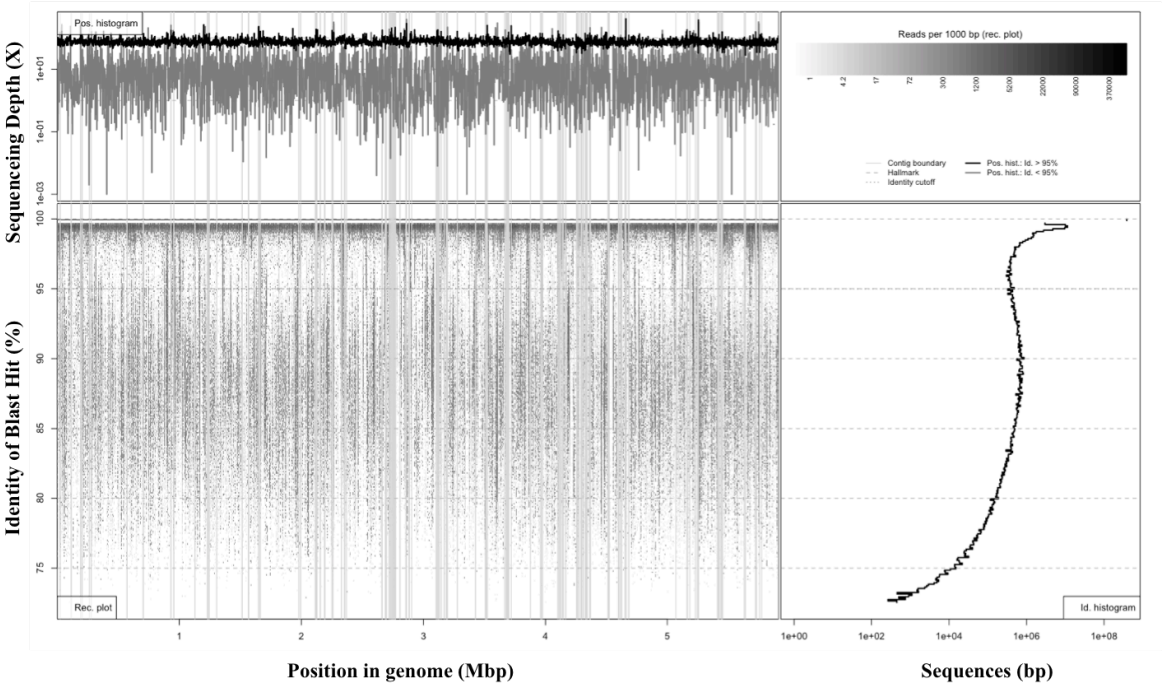
Fragment Recruitment of Bin 04 vs. AK-T4 sample metagenomic dataset



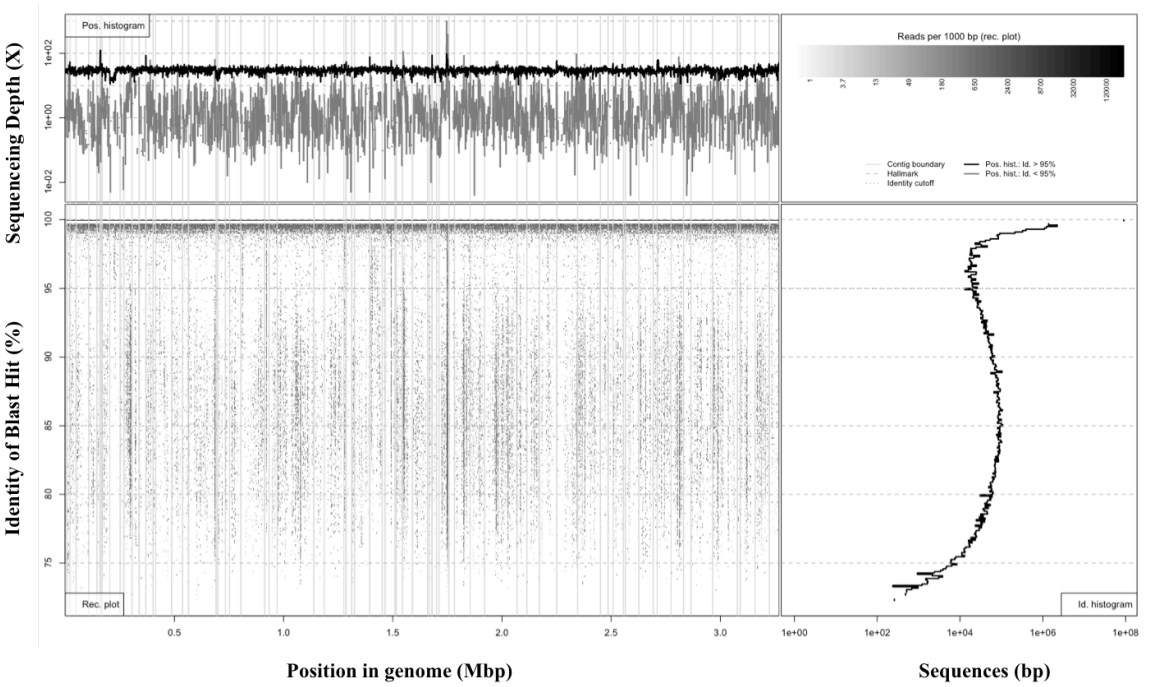
Fragment Recruitment of Bin 06 vs. AK-C2 sample metagenomic dataset



Fragment Recruitment of Bin 07 vs. AK-T4 sample metagenomic dataset



Fragment Recruitment of Bin 08 vs. AK-C4 sample metagenomic dataset



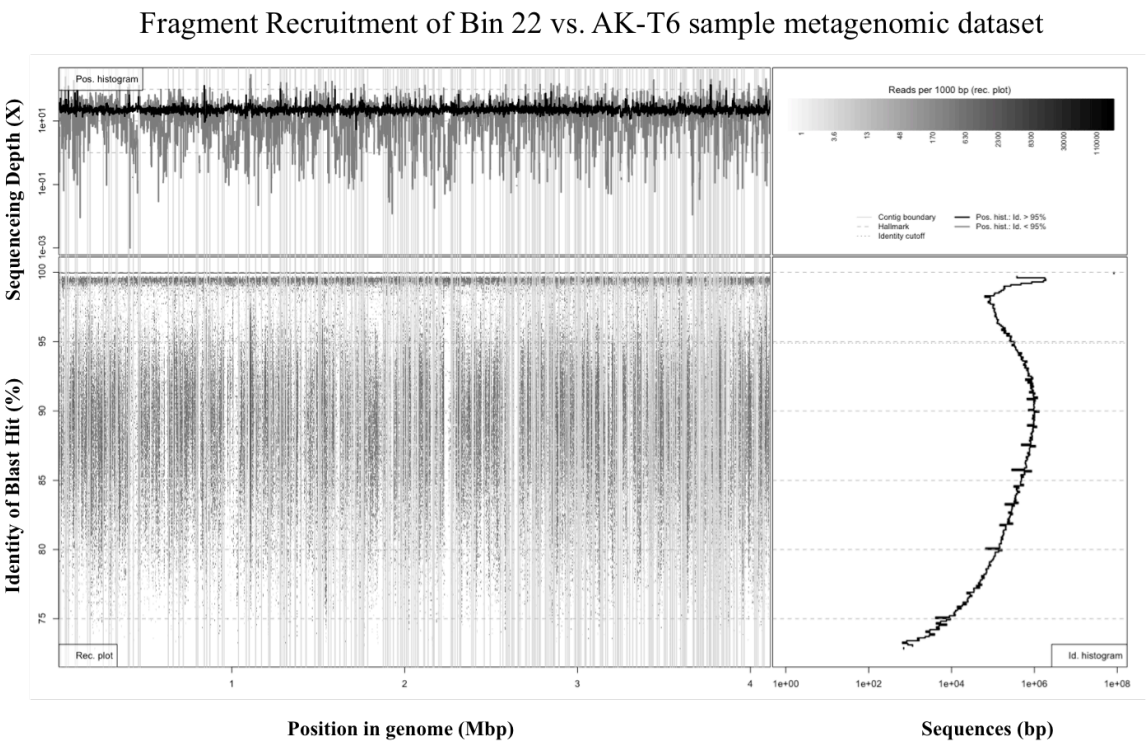
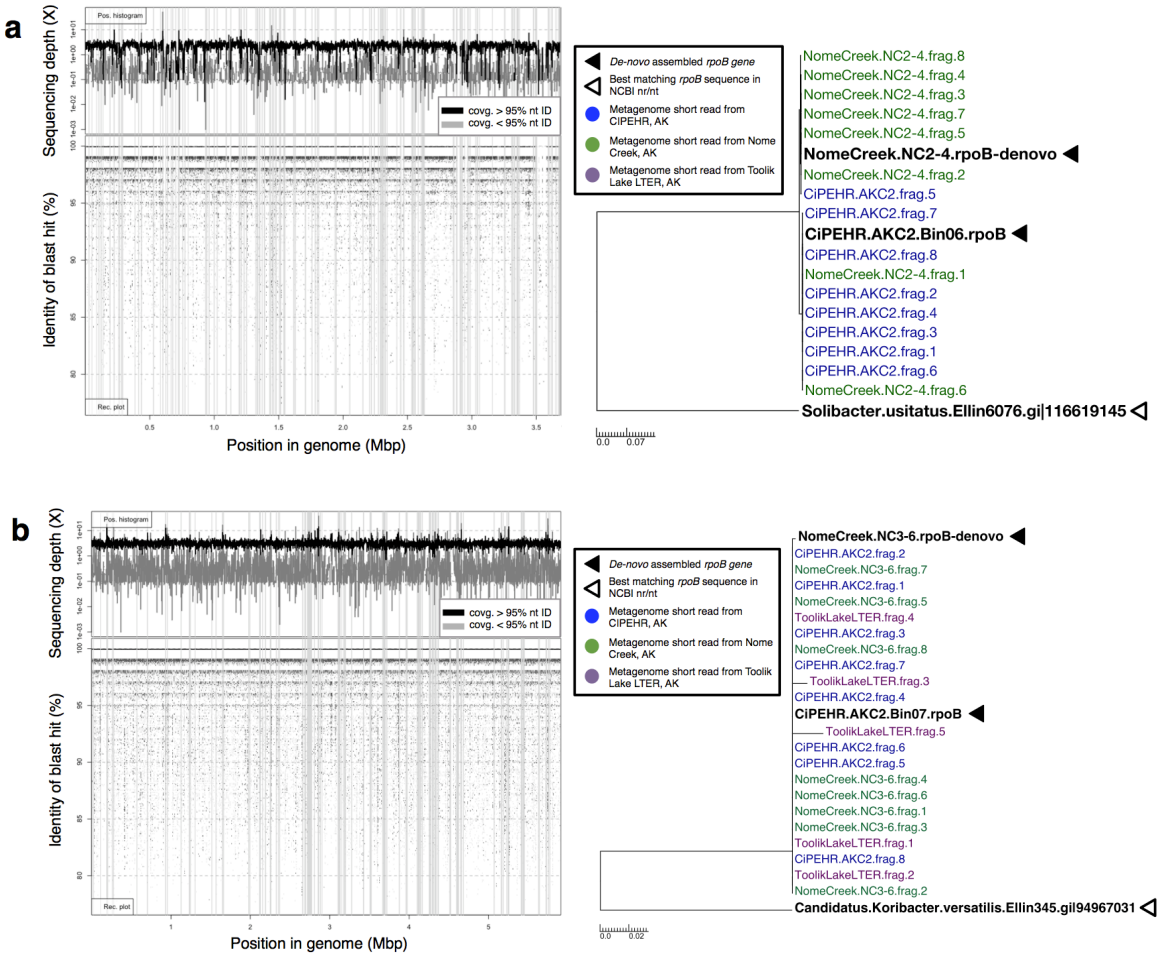


Figure C-7: Fragment recruitment plots. Diagrams expressing coverage of all contigs of several population bins assembled from AK metagenomes. All metagenomic short reads from select sample metagenomes (usually the sample in which the organism had the highest representation) were searched against each bin in a megablast search. The position histogram (top left of each) displays the coverage of each base position, determined by a 1000bp window average. An even coverage across the entire contig from matching reads that are high identity (>95%) as well as an absence of redundant positions is indicative of a high quality contig from a single population. The recruitment plot (bottom left of each) shows where individual metagenomic reads matched to the population bin and the identity (%) of the match. The ID histogram (bottom right) displays the total number of short read-derived base-positions at given percent identities. Note that in all cases shown, a sequence-discrete population represented by

reads showing high nucleotide identity to the reference population genome sequence (typically >95% nucleotide identity) and even coverage across the length of the reference sequence are obvious. In some cases such as Bin 22 against AK-T6 closely related (i.e., showing 80-95% nucleotide identity to the reference sequence) population(s) are also present (co-occurring) in the sample while in cases such as Bin 04 against AK-T4, no co-occurring relatives are present (or those are of very low abundance to be robustly detected by metagenomics)

.



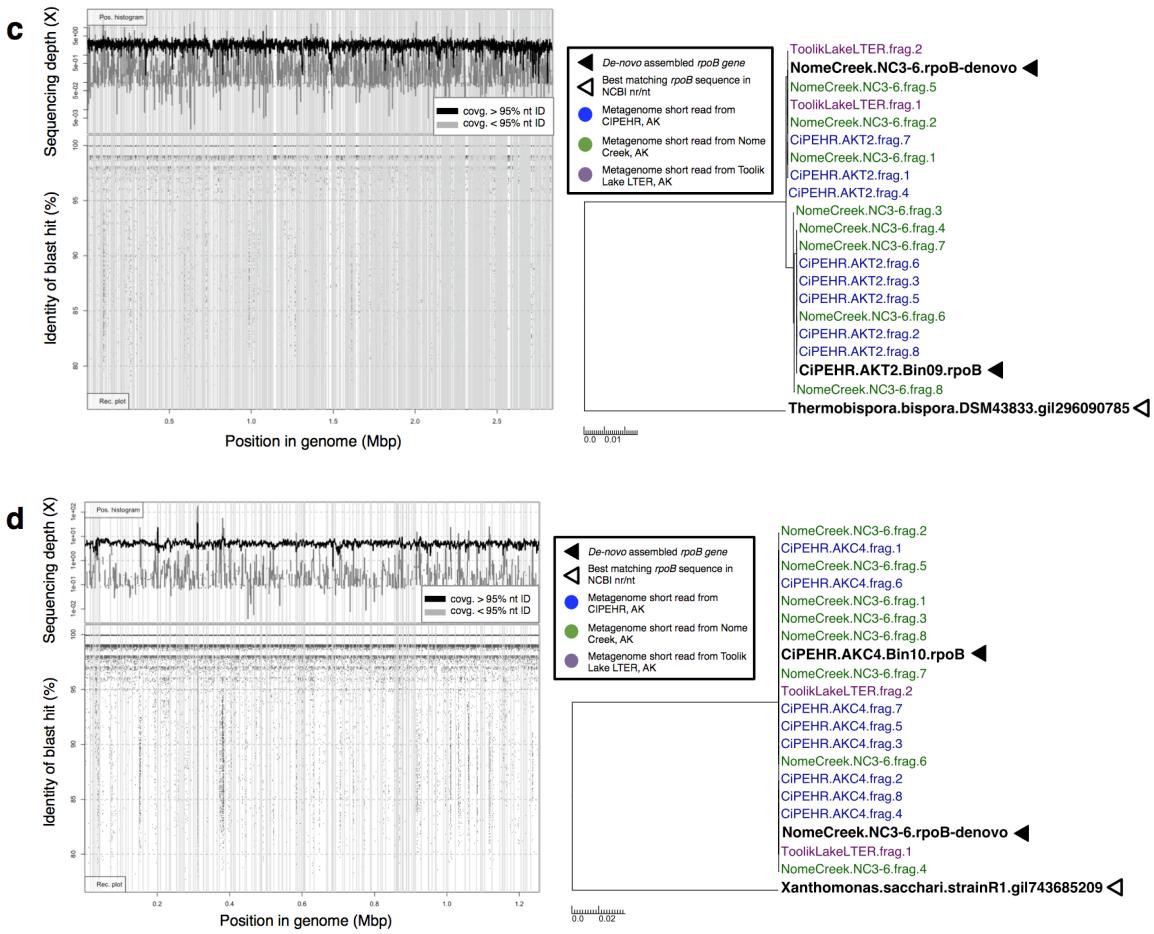


Figure C-8: Fragment recruitment plots and phylogenetic trees of *rpoB* gene sequences of four geographically dispersed population bins. Population bins assembled from CiPEHR metagenomic datasets were searched against publically available metagenomes representing distant locations for fragment recruitment analysis in the same manner as discussed in Figure C-7. Diagrams express the coverage of concatenated contigs (from left to right; contigs separated by grey vertical bars) in metagenomes from Taş et al. 2014 (Nome Creek, AK; >200km distance from CiPEHR). Phylogenetic trees constructed from *rpoB* gene sequences assembled, *de-novo,* from metagenomes at both locations independently (in boldface) as well as unassembled *rpoB*-

encoding short reads from the same metagenome. Trees also include a reference *rpoB* gene sequence from NCBI (outgroup) as well as unassembled reads from a soil metagenome from Fierer et al. 2012 (Toolik Lake LTER, AK; ~530km distance from CiPEHR). The trees reveal the high relatedness among organisms present at all three sites (e.g., reads do not cluster by site but are intermixed). Diagrams represent (**a**) Sample NC2-4 (Burned 10-20cm) vs Bin 06 assembly, (**b**) Sample NC3-6 (Control 50-60cm) vs Bin 07 assembly, (**c**) Sample NC3-6 vs Bin 09 assembly, (**d**) and Sample NC3-6 vs Bin 10 assembly.

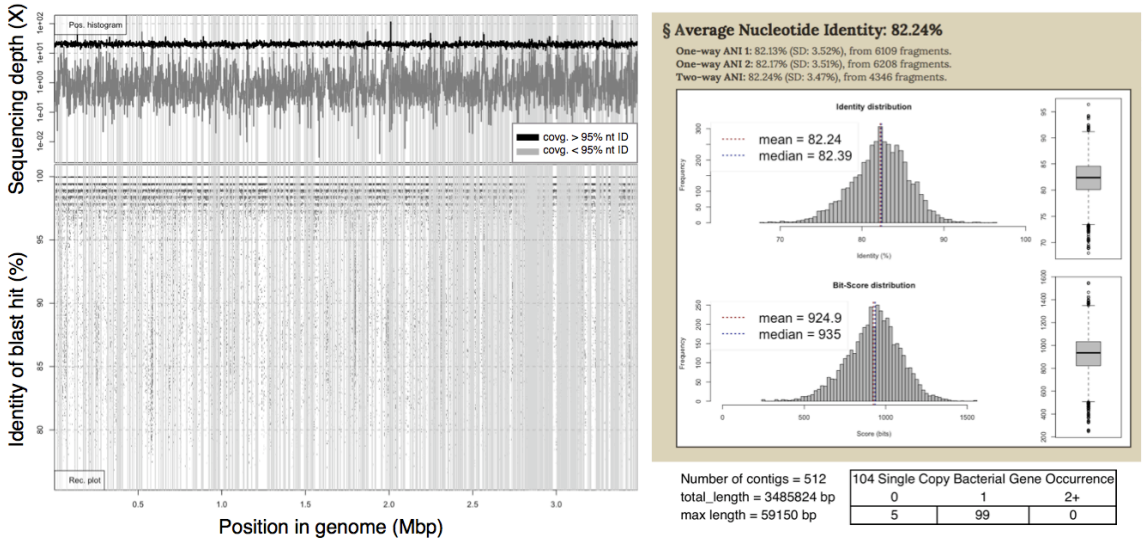


Figure C-9: The acidobacterial population recovered from the SPRUCE metagenome. Graphs show the acidobacterial fragment recruitment plot of SPRUCE soil metagenome assembly (left) average nucleotide identity comparison to bin 11 (right top), basic assembly statistics, and assessment of bin completeness and contamination (bottom right table). Population bin assembled from SPRUCE soil metagenomes were searched (with megablast) against the

dataset from which it was derived (NCBI SRA entry SRR1157608; 75-100 cm soil depth) and a fragment recruitment plot was generated in the same manner as discussed in Figure C-5. Average nucleotide identity (ANI) comparison was made between this assembled bin and the closest matching assembly from CiPEHR metagenomes, bin 11, using web-based ANI calculation tools available at http://enve-omics.ce.gatech.edu/ani/. Presence of 104 single copy bacterial genes, which is used in the assessment of assembly completion and contamination, was performed using CheckM.

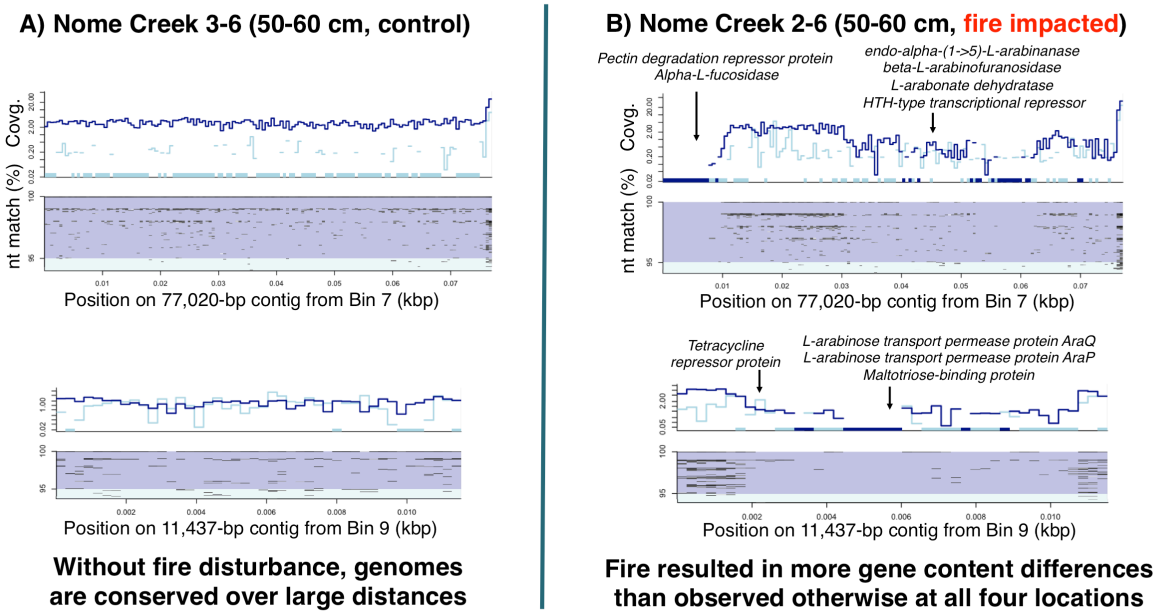


Figure C-10: Contigs from two populations conserved between CiPEHR and Nome Creek control soils, but with gene loss resulting from fire event at Nome Creek. Contigs from population bins assembled from CiPEHR metagenomic datasets were searched against publically available metagenomes representing A) Nome Creek control 50-60cm depth and B) Nome Creek fire

impacted 50-60cm depth) for fragment recruitment analysis in the same manner as discussed in Figure C-7. Gene content difference was 2-3% of the entire genome assembly in the case of bins 7 and 9. Contigs shown above displayed higher than usual gene loss and thus, are displayed for contrast.

Site	Sample ID	Bulk density	%N	%C	pH	LP1 (mg C g dry soil^{-1})	LP2 (mg C g dry soil^{-1})	RP (mg C g dry soil^{-1})	TOC (mg C/L)
CiPEHR	AK-C1	0.14	2.00	37.68	5.05	115.90	37.33	423.70	1195.00
CiPEHR	AK-C2	0.28	1.56	35.44	4.94	123.50	50.87	366.10	1356.50
CiPEHR	AK-C3	0.18	0.40	38.66	4.80	154.85	55.80	394.50	1094.00
CiPEHR	AK-C4	0.33	1.20	18.71	4.86	141.74	106.80	311.60	640.00
CiPEHR	AK-C5	0.21	1.52	31.27	4.86	36.86	38.00	193.80	768.00
CiPEHR	AK-C6	0.63	1.81	36.11	5.23	150.10	73.70	415.00	1158.50
CiPEHR	AK-T1	0.82	1.75	33.64	4.73	191.90	121.80	380.30	809.50
CiPEHR	AK-T2	0.17	2.95	27.60	4.86	83.41	59.20	272.10	735.00
CiPEHR	AK-T3	0.21	1.55	32.39	4.51	37.70	70.05	284.20	646.00
CiPEHR	AK-T4	0.21	1.18	38.54	4.63	75.32	69.34	349.90	889.50
CiPEHR	AK-T5	0.21	1.05	34.77	4.44	51.05	64.24	311.30	841.00
CiPEHR	AK-T6	0.20	1.13	34.74	4.96	96.52	76.82	298.30	1071.50
KFFL	OK-3C	N/A	4.98	0.81	N/A	6.41	2.44	3.80	56.00
KFFL	OK-6C	N/A	5.89	0.42	N/A	2.18	1.58	2.80	23.30
KFFL	OK-19C	N/A	5.41	0.55	N/A	1.69	1.20	2.90	17.95
KFFL	OK-23C	N/A	4.28	0.89	N/A	3.16	2.61	5.60	35.65
KFFL	OK-1H	N/A	5.26	0.48	N/A	3.03	2.51	4.60	34.20
KFFL	OK-10H	N/A	4.60	0.84	N/A	3.12	4.33	4.70	45.15
KFFL	OK-16H	N/A	3.64	1.16	N/A	2.73	3.32	5.30	36.80
KFFL	OK-20H	N/A	4.03	1.15	N/A	4.30	5.54	7.90	59.80

Table C-1: Soil chemical and physical measurements for CiPEHR and KFFL soil samples. Method descriptions for each type of measurement can be found in Xue et al. 2016.

Site Name	Approximate Location	Provided GPS Coordinates	Biom Type	Soil Sampling Date	Sampling Depth(s)	Applicable Reference
Carbon in Permafrost Experimental Heating Research Site	Denali National Park, Alaska, USA	63º52'59''N, 149º13'32''W	Arctic Tundra	May, 2010	15-25 cm	-
Kessler Farm Field Laboratory Research Site	McClain County, Oklahoma, USA	34°58'54"N, 97°31'14"W	Temperate Grassland	October, 2010	0-15 cm	-
Toolik Late Long Term Ecological Research Station	Toolik Lake Region, Alaska, USA	68°38'N, 149°43'W	Arctic Tundra	*Not Provided*	0-5 cm	Fierer et al., 2012
Bonanza Creek Long-Term Ecological Research Program	near Fairbanks, Alaska, USA	64°42'N 148°18'W	Arctic Tundra	Summer 2009	30-35 cm	Hultman et al., 2015
Nome Creek Area	~100 km NE of Fairbanks, Alaska, USA	65°20'42.7"N, 146°55'08.4"W	Arctic Tundra	September, 2011	10-20, 50-60, and 90-100 cm	Taş et al., 2014
Agricultural Location in Urbana, IL, USA	Urbana, Illinois, USA	*Not provided*	Temperate Agriculture	November, 2011	0-30 cm	Orellana et al., 2014
Inner Mongolia Grassland Ecosystem Research Station	Inner Mongolia, China	43°38´N, 116°42'E	Temperate Steppe	August 22, 2010	0-10 cm	Zhang et al., 2016

Table C-2: Datasets (including pertinent information) used to compare community characteristics (taxa, populations, functional genes) shared within Alaskan tundra soils.

Sample ID	PCR Amplicon Barcode Identifier	Reads assigning to sample	Numb of Reads Merged	Numb of Merged Seqs. GTE 190bp	Seqs. Annotating to Bacterial Phylum
AK_C1	CCTGGTCTTACGGC	94,412	85,707	27,949	27,181
AK_C2	CGACCGATAGGGAC	82,937	76,795	32,420	32,154
AK_C3	CACGTGAGGAACGC	83,361	74,946	28,009	27,413
AK_C4	CGGTCTAGGTCTAC	78,368	73,475	31,592	30,685
AK_C5	CATCGAATCGAGTC	91,600	85,431	37,176	36,424
AK_C6	CGGATGCAGGATGC	94,084	83,654	31,480	30,879
AK_T1	CCAAGTCGAATACC	85,791	79,784	30,704	29,996
AK_T2	CGGCGAACTGAAGC	95,337	84,762	32,100	31,483
AK_T3	CTGAATCGAAGCTC	80,734	74,066	28,361	27,451
AK_T4	CTCAGGACGTATCC	83,954	78,175	31,287	30,566
AK_T5	CCTAGCAGTATGAC	81,786	76,076	29,970	29,379
AK_T6	CCCACTTGAGAGTC	92,911	86,542	34,376	33,751
OK_3C	CCTTTAGCGCTGGC	93,508	83,989	24,929	24,234
OK_6C	CTTAGACTCGGAAC	75,412	70,468	25,164	24,461
OK_19C	CCCAAACTCGTCGC	82,919	75,133	22,272	21,637
OK_23C	CGTGCACGATAATC	68,301	64,619	24,318	23,644
OK_1H	CGGTACTGTACCAC	76,396	70,865	24,170	23,373
OK_10H	CAGTACCTAAGTGC	88,088	78,376	24,185	23,367
OK_16H	CGTCTCTGAAAGAC	74,939	69,332	24,020	23,463

Table C-3: Information pertaining to sample-specific barcode identifiers, the merge-ability, sequence quality, and number of annotation made for 16S rRNA gene amplicon sequences representative of each soil sample. For sample 'OK_20H', PCR amplicon data were not available. Therefore, it was excluded from this portion of the analyses.

Sample ID	Total Number of Raw Seqs.	Total Sequencing depth (bp)	Number of Merged Seqs.	Merged Seqs. GTE 100bp After Trimming	Seqs. Annotated to a SwissProtKB Gene	Perc. of Query Seqs. Annotated
AK_C2	140,384,743	42,115,422,900	88,938,535	81,911,488	11,272,986	13.76%
AK_C3	127,668,458	38,300,537,400	55,494,410	50,258,776	7,133,996	14.19%
AK_C4	123,994,800	37,198,440,000	43,931,094	41,215,414	7,007,438	17.00%
AK_C5	136,918,270	41,075,481,000	77,972,852	74,233,948	11,093,641	14.94%
AK_C6	145,020,548	43,506,164,400	88,303,693	79,740,381	9,816,350	12.31%
AK_T1	88,621,343	26,586,402,900	43,063,960	41,082,354	6,082,884	14.81%
AK_T2	115,149,822	34,544,946,600	52,812,203	47,829,930	6,689,572	13.99%
AK_T3	129,728,436	38,918,530,800	70,847,511	65,415,616	9,846,729	15.05%
AK_T4	144,973,423	43,492,026,900	110,018,815	101,786,723	12,711,784	12.49%
AK_T5	166,411,844	49,923,553,200	110,104,603	100,970,305	12,602,459	12.48%
AK_T6	149,370,367	44,811,110,100	109,824,451	99,923,813	12,167,769	12.18%
OK_3C	120,253,701	36,076,110,300	52,477,090	48,578,565	8,036,148	16.54%
OK_6C	109,642,544	32,892,763,200	64,581,913	20,384,797	3,266,908	16.03%
OK_19C	132,777,103	39,833,130,900	65,725,755	60,307,185	9,293,533	15.41%
OK_23C	116,865,558	35,059,667,400	57,197,196	52,759,317	8,589,159	16.28%
OK_1H	146,224,604	43,867,381,200	70,075,328	64,873,555	10,775,911	16.61%
OK_10H	135,467,362	40,640,208,600	63,057,030	58,630,367	9,609,617	16.39%
OK_16H	141,761,229	42,528,368,700	38,561,786	31,704,520	4,452,998	14.05%
OK_20H	123,179,662	36,953,898,600	71,373,159	65,363,171	10,388,444	15.89%

Table C-4: Information pertaining to the merge-ability, sequence quality, and number/percent of annotation made for shotgun-metagenome datasets representing each soil sample.

Metagenome Dataset	$kappa	Estimated % Coverage	Depth to reach 95% coverage (bp)	Actual Depth (bp)
CiPEHR AK-C2	0.750	0.804	8.86E+10	1.51E+10
CiPEHR AK-C3	0.700	0.764	6.18E+10	9.44E+09
CiPEHR AK-C4	0.666	0.735	5.04E+10	8.41E+09
CiPEHR AK-C5	0.813	0.855	4.30E+10	1.47E+10
CiPEHR AK-C6	0.839	0.876	3.35E+10	1.31E+10
CiPEHR AK-T1	0.784	0.831	3.12E+10	8.15E+09
CiPEHR AK-T2	0.717	0.777	4.60E+10	8.70E+09
CiPEHR AK-T3	0.729	0.788	7.05E+10	1.21E+10
CiPEHR AK-T4	0.754	0.807	7.49E+10	1.68E+10
CiPEHR AK-T5	0.840	0.876	4.39E+10	1.67E+10
CiPEHR AK-T6	0.820	0.861	4.63E+10	1.63E+10
KFFL OK-3C	0.293	0.394	1.47E+12	8.94E+09
KFFL OK-6C	0.224	0.322	1.46E+11	3.71E+09
KFFL OK-19C	0.263	0.364	4.54E+11	1.10E+10
KFFL OK-23C	0.417	0.516	2.12E+11	9.59E+09
KFFL OK-1H	0.369	0.470	4.82E+11	1.20E+10
KFFL OK-10H	0.223	0.321	2.87E+11	1.08E+10
KFFL OK-16H	0.215	0.313	3.11E+11	5.44E+09
KFFL OK-20H	0.541	0.628	2.35E+11	1.16E+10
Comprehensive CiPEHR AK	0.893	0.918	2.13E+11	1.39E+11
Comprehensive KFFL OK	0.485	0.579	1.84E+12	6.22E+10
NC1-4 (Fire A 10-20cm)	0.506	0.597	5.34E+10	5.39E+09
NC1-6 (Fire A 50-60cm)	0.646	0.718	3.87E+10	6.38E+09
NC1-12 (Fire A 90-100cm)	0.758	0.811	1.94E+10	5.81E+09
NC2-4 (Fire B 10-20cm)	0.375	0.476	1.07E+11	3.15E+09
NC2-6 (Fire B 50-60cm)	0.589	0.669	6.08E+10	8.01E+09
NC2-12 (Fire B 90-100cm)	0.523	0.612	1.07E+11	6.93E+09
NC3-4 (Control A 10-20cm)	0.427	0.525	5.89E+10	5.79E+09
NC3-6 (Control A 50-60cm)	0.678	0.745	2.50E+10	6.14E+09
NC3-12 (Control A 90-100cm)	0.610	0.688	4.37E+10	5.69E+09
NC4-4 (Control B 10-20cm)	0.369	0.470	7.00E+10	6.01E+09
NC4-6 (Control B 50-60cm)	0.639	0.712	3.40E+10	5.94E+09
NC4-12 (Control B 90-100cm)	0.769	0.820	1.54E+10	5.48E+09
Bonanze Creek Active Layer A	0.164	0.254	2.42E+10	1.15E+09
Bonanze Creek Active Layer B	0.207	0.303	4.49E+10	1.93E+09
Urbana, IL (Agriculture)	0.233	0.331	2.81E+11	1.25E+10
Steppe - Inner Mongolia, China	0.189	0.283	2.15E+11	4.65E+09

Table C-5: Soil microbial community complexity for estimations as determined by Nonpareil. Nonpareil is a statistical tool that uses read redundancy to estimate dataset complexity and the amount of sequencing effort needed to achieve a desired level of åcoverage, given for each sample, as well as ‘comprehensive datasets’ containing all reads from either site.

Taxon	Domain	Taxon Rank	OK_3C	OK_6C	OK_19C	OK_23C	OK_1H	OK_6H	OK_16H	OK-Mean
Nitrospira	*Bacteria*	Genus	0.907%	0.608%	0.893%	0.493%	0.832%	0.538%	0.408%	0.668%
Nitrospiraceae	*Bacteria*	Family	1.091%	0.838%	1.069%	0.633%	1.129%	0.765%	0.529%	0.865%
Thermodesulfovibrionaceae	*Bacteria*	Family	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Nitrososphaeraceae	*Archaea*	Family	0.469%	0.385%	0.162%	0.136%	0.550%	0.385%	0.445%	0.362%
Methanobacteriaceae	*Archaea*	Family	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Nitrosomonas	*Bacteria*	Genus	0.000%	0.000%	0.004%	0.000%	0.000%	0.004%	0.000%	0.001%
Chromatiales	*Bacteria*	Order	0.012%	0.056%	0.058%	0.070%	0.058%	0.004%	0.033%	0.042%
Rhodospirillaceae	*Bacteria*	Family	1.697%	1.534%	1.625%	1.793%	1.837%	2.171%	1.982%	1.805%
Chloracidobacterium	*Bacteria*	Class	2.463%	2.742%	3.367%	2.665%	2.971%	1.943%	3.014%	2.738%

Taxon	AK_C1	AK_C2	AK_C3	AK_C4	AK_C5	AK_C6	AK_T1	AK_T2	AK_T3	AK_T4	AK_T5	AK_T6	AK-Mean
Nitrospira	0.061%	0.000%	0.000%	0.000%	0.000%	0.003%	0.029%	0.000%	0.000%	0.000%	0.000%	0.000%	0.008%
Nitrospiraceae	0.061%	0.000%	0.000%	0.000%	0.000%	0.003%	0.033%	0.000%	0.000%	0.000%	0.000%	0.000%	0.008%
Thermodesulfovibrionaceae	0.007%	0.318%	0.029%	0.013%	0.016%	0.883%	0.049%	0.302%	0.219%	0.400%	0.397%	0.099%	0.228%
Nitrososphaeraceae	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Methanobacteriaceae	0.011%	0.000%	0.104%	0.006%	0.048%	0.010%	0.000%	0.037%	0.004%	0.109%	0.037%	0.017%	0.032%
Nitrosomonas	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Chromatiales	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Rhodospirillaceae	0.827%	0.287%	0.143%	0.237%	0.038%	0.102%	0.244%	0.206%	0.028%	0.067%	0.060%	0.084%	0.194%
Chloracidobacterium	0.000%	0.022%	0.000%	0.000%	0.000%	0.000%	0.000%	0.009%	0.067%	0.003%	0.000%	0.000%	0.008%

Table C-6: The Abundance of *Bacteria* and *Archaea* taxonomic groups of interest with means specific to each study site. Data shown in table was determined by 16S rRNA gene amplicons of the V4 region. Oklahoma sample 20H is omitted from this analysis due to technical difficulties during library preparation.

Alaska Soil Samples

Phylum	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6	Mean	Std. Dev.
Acidobacteria	29.96%	51.80%	49.76%	57.89%	64.68%	62.00%	47.67%	49.08%	28.15%	54.78%	59.52%	58.81%	51.17%	11.62%
Proteobacteria	33.99%	12.96%	13.21%	14.73%	12.45%	13.31%	16.28%	16.56%	17.52%	14.46%	11.24%	9.77%	15.54%	6.22%
Verrucomicrobia	9.70%	22.50%	20.70%	9.52%	9.18%	13.87%	19.11%	18.00%	13.16%	14.29%	14.09%	16.75%	15.07%	4.43%
Actinobacteria	14.77%	4.12%	7.74%	9.57%	8.31%	4.58%	8.21%	8.53%	11.26%	8.61%	9.18%	9.43%	8.69%	2.77%
Bacteroidetes	1.90%	3.20%	1.64%	1.34%	0.75%	0.57%	1.63%	1.50%	17.86%	1.01%	0.75%	0.72%	2.74%	4.82%
Bacteria;Other	2.63%	0.73%	1.97%	2.85%	1.90%	1.85%	2.22%	1.84%	3.18%	2.12%	1.84%	1.72%	2.07%	0.62%
Planctomycetes	2.92%	1.27%	0.90%	0.76%	0.36%	0.63%	2.23%	0.68%	0.44%	1.24%	0.51%	0.53%	1.04%	0.79%
AD3	1.37%	0.80%	0.81%	0.74%	1.07%	1.31%	0.94%	1.12%	1.41%	0.86%	0.90%	0.77%	1.01%	0.24%
Gemmatimonadetes	0.29%	0.35%	0.38%	1.33%	0.31%	0.15%	0.29%	0.44%	1.53%	0.28%	0.26%	0.19%	0.48%	0.45%
Chlorobi	0.05%	0.51%	1.11%	0.18%	0.24%	0.19%	0.08%	0.46%	1.54%	0.58%	0.40%	0.37%	0.48%	0.44%
Chloroflexi	0.44%	0.19%	0.48%	0.41%	0.07%	0.10%	0.17%	0.40%	2.40%	0.05%	0.07%	0.16%	0.41%	0.65%
Elusimicrobia	0.57%	0.45%	0.50%	0.33%	0.15%	0.22%	0.30%	0.40%	0.34%	0.58%	0.20%	0.28%	0.36%	0.14%
Nitrospirae	0.15%	0.32%	0.03%	0.01%	0.02%	0.89%	0.08%	0.30%	0.22%	0.40%	0.40%	0.10%	0.24%	0.25%
Armatimonadetes	0.26%	0.25%	0.14%	0.08%	0.07%	0.06%	0.16%	0.10%	0.16%	0.11%	0.13%	0.04%	0.13%	0.07%
Cyanobacteria	0.33%	0.14%	0.07%	0.06%	0.09%	0.04%	0.10%	0.07%	0.10%	0.12%	0.03%	0.05%	0.10%	0.08%
Firmicutes	0.08%	0.08%	0.12%	0.03%	0.03%	0.05%	0.07%	0.22%	0.18%	0.07%	0.09%	0.07%	0.09%	0.06%
FCPU426	0.19%	0.08%	0.10%	0.08%	0.13%	0.02%	0.07%	0.07%	0.02%	0.06%	0.11%	0.06%	0.08%	0.05%
WPS-2	0.16%	0.05%	0.01%	0.00%	0.02%	0.02%	0.21%	0.02%	0.12%	0.04%	0.02%	0.02%	0.06%	0.07%

Oklahoma Soil Samples

Phylum	3C	6C	19C	23C	1H	6H	16H	Mean	Std. Dev.
Proteobacteria	27.96%	23.03%	33.42%	24.04%	27.34%	34.53%	33.77%	29.16%	4.42%
Verrucomicrobia	16.88%	27.36%	18.73%	33.72%	19.83%	14.85%	14.85%	20.89%	6.56%
Acidobacteria	19.95%	24.10%	18.97%	19.84%	21.37%	18.40%	18.55%	20.17%	1.86%
Actinobacteria	9.41%	6.04%	7.16%	6.44%	9.23%	12.46%	13.87%	9.23%	2.78%
Bacteroidetes	4.03%	3.68%	5.91%	2.46%	3.57%	3.82%	4.75%	4.03%	0.99%
Planctomycetes	3.83%	3.19%	3.89%	3.07%	3.41%	2.74%	3.22%	3.34%	0.38%
Firmicutes	6.41%	2.21%	1.13%	1.63%	3.14%	1.97%	0.91%	2.49%	1.74%
Bacteria;Other	2.21%	2.32%	2.63%	2.59%	2.69%	2.97%	1.81%	2.46%	0.35%
Gemmatimonadetes	2.09%	1.73%	1.79%	1.33%	2.06%	2.16%	2.05%	1.89%	0.27%
Chloroflexi	2.03%	1.72%	1.85%	1.20%	1.81%	2.06%	1.97%	1.80%	0.27%
Nitrospirae	1.53%	1.14%	1.37%	1.05%	1.46%	1.11%	0.82%	1.21%	0.23%
Cyanobacteria	0.58%	0.47%	0.70%	0.56%	0.82%	0.60%	1.32%	0.72%	0.27%
WS3	0.59%	0.91%	0.74%	0.44%	0.48%	0.55%	0.44%	0.59%	0.16%
Elusimicrobia	0.48%	0.60%	0.46%	0.44%	0.56%	0.39%	0.27%	0.46%	0.10%
Crenarchaeota	0.47%	0.41%	0.16%	0.14%	0.55%	0.38%	0.45%	0.37%	0.14%
AD3	0.40%	0.23%	0.19%	0.30%	0.51%	0.24%	0.18%	0.29%	0.11%
Armatimonadetes	0.17%	0.23%	0.26%	0.14%	0.22%	0.18%	0.24%	0.20%	0.04%
TM6	0.24%	0.15%	0.11%	0.13%	0.14%	0.17%	0.11%	0.15%	0.04%
OP3	0.23%	0.09%	0.08%	0.14%	0.19%	0.09%	0.08%	0.13%	0.06%
Chlamydiae	0.12%	0.10%	0.11%	0.08%	0.07%	0.08%	0.05%	0.09%	0.02%

Table C-7: Abundance of Bacterial and Archaeal phyla in (top) Alaskan tundra CiPHER site soils, and (bottom) Oklahoma prairie KFFL site soils. Means and standard deviations specific to each site are displayed. Data shown in table was determined by 16S rDNA PCR amplicons of the V4 region. Phyla are in order from highest to lowest mean for each site. Phyla with a mean abundance of less than 0.5% are omitted from each table. Note that Oklahoma sample 20H is omitted from this analysis due to technical difficulties during library preparation.

Selected Genes	OK_3C	OK_6C	OK_19C	OK_23C	OK_1H	OK_10H	OK_16H	OK_20H	AK_C2	AK_C3	AK_C4	AK_C5	AK_C6	AK_T1	AK_T2	AK_T3	AK_T4	AK_T5	AK_T6
Magnesium-chelatase subunit (*bchD*)	13.216	12.800	12.621	13.284	7.335	36.087	4.591	9.254	0.710	0.701	1.284	1.532	1.222	0.329	0.448	1.320	1.416	0.873	0.740
Magnesium-protoporphyrin IX monomethyl ester cyclase (*acsF*)	3.018	6.512	3.155	0.866	0.466	2.987	0.612	2.690	0.355	0.841	0.571	0.000	0.407	0.000	0.000	0.305	1.101	0.159	0.329
Anaerobic magnesium-protoporphyrin IX monomethyl ester cyclase (*bchE*)	51.095	41.770	37.955	65.650	48.899	39.073	41.017	46.054	18.983	19.484	17.267	23.797	17.929	20.056	28.103	23.967	17.936	16.743	21.861
Cytochrome c2 proteins (*cycA*)	28.201	33.461	26.633	27.627	26.662	24.265	30.304	28.945	12.774	13.317	19.551	13.882	8.965	16.111	12.856	17.671	14.475	16.108	9.040
Nitrogenase iron protein (*nifH*)	0.728	2.246	0.650	0.289	1.164	0.373	0.612	0.861	59.168	115.083	60.079	53.274	71.819	54.744	92.383	82.464	136.488	66.098	95.334
Nitrite Reductase (*nirK*)	18.939	19.762	14.941	14.439	16.300	16.550	13.774	16.355	6.210	3.364	14.984	8.293	5.297	7.069	5.531	5.586	3.855	6.427	5.506
Nitrite Reductase (*nirS*)	2.185	3.818	1.578	0.481	1.746	1.618	2.755	4.734	0.089	0.000	0.714	0.000	0.204	0.000	0.299	0.102	0.000	0.079	0.000
Nitrous-oxide reductase (*nosZ*)	55.257	36.155	69.228	96.935	69.157	61.223	99.482	50.788	19.516	12.616	34.107	19.561	9.067	28.769	24.815	52.911	7.631	11.109	14.793
Archaeal ammonia monooxygenase (*amoA*)	9.157	6.962	6.496	5.198	8.383	7.840	10.101	11.513	1.331	0.701	2.854	0.361	0.815	2.959	1.495	0.203	1.495	0.555	0.411
Sulfate adenylyltransferase subunit 1 (*cysD*)	424.262	431.170	352.453	360.112	369.419	382.273	398.848	398.234	468.731	492.431	523.301	467.836	525.654	414.113	515.878	519.767	559.638	487.048	589.755
Sulfate adenylyltransferase subunit 2 (*cysN*)	304.174	302.942	275.615	274.632	280.703	295.166	296.305	300.639	413.200	438.464	453.661	386.257	444.259	266.321	374.762	454.059	459.416	404.604	471.820
Sulfur Oxidation (*soxA*)	55.049	65.125	52.246	30.515	62.288	50.771	51.731	58.428	12.153	8.551	21.406	14.603	8.353	13.974	14.949	20.413	11.879	7.935	6.328
Methylmalonyl-CoA mutase small subunit (*mutA)*	78.359	96.564	65.981	59.200	84.176	78.645	63.975	81.347	124.989	121.811	101.606	101.049	116.133	130.366	110.022	173.459	119.259	105.932	105.607
Phosphate acetyltransferase (*pta*)	2.497	2.246	3.990	3.369	3.027	3.111	2.755	2.905	7.274	5.186	9.561	9.285	8.150	14.631	11.809	5.890	8.889	8.014	13.232
Acetate kinase 1,2 (*ackA1, ackA2)*	0.728	0.674	1.206	0.866	0.582	1.120	0.000	0.646	1.153	0.701	0.856	1.082	0.611	2.795	1.644	1.320	1.259	0.793	0.740
Carbon monoxide dehydrogenase/acetyl-CoA synthase subunit alpha/beta	0.624	1.797	0.928	0.096	0.116	0.622	1.530	0.861	151.867	198.626	109.312	168.024	241.332	53.922	114.656	148.069	130.666	136.719	213.433
5-methyltetrahydrofolate:corrinoid/iron-sulfur co-methyltransferase (*acsE*)	1.457	2.021	2.227	2.021	2.096	1.867	0.612	0.968	4.968	3.504	10.703	6.220	3.158	5.918	7.026	24.170	3.855	4.364	3.205
Putative phosphonates transport system permease (*phnE, phnD*)	48.805	40.871	42.595	32.344	41.331	43.802	33.059	52.832	13.750	10.513	21.977	10.727	6.622	13.645	10.913	26.202	11.643	9.046	6.246
Phosphate-import permease (*phnE, phnD*)	11.135	11.902	8.630	5.679	7.218	10.951	8.571	10.007	2.040	1.402	2.141	1.713	1.630	1.973	2.242	2.031	1.809	1.349	1.315
Inorganic pyrophosphatase (*ppa*)	328.421	368.965	295.845	293.403	271.039	311.965	296.305	326.248	137.586	158.957	167.536	147.923	136.507	183.137	186.559	188.997	214.132	141.401	158.780
Lysophospholipase (*nte1*)	37.671	35.706	38.976	31.285	38.886	38.576	45.609	41.319	17.209	20.045	27.257	23.978	16.707	19.234	17.938	21.936	30.366	20.155	17.587
Methyl-coenzyme M reductase subunit alpha (*mcrA*)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.177	2.523	2.854	2.344	0.611	0.000	0.299	5.484	1.337	1.666	0.740
Methyl-coenzyme M reductase subunit beta (*mcrB*)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.981	0.714	1.442	0.407	0.164	0.747	5.484	1.888	1.825	0.247
Acetyl-CoA decarbonylase/synthase complex subunit alpha (*cdhA*)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.976	1.682	0.999	2.434	2.445	0.000	2.691	7.820	2.360	2.777	2.876
Acetyl-CoA decarbonylase/synthase complex subunit beta (*cdhC*)	0.000	0.225	0.278	0.096	0.000	0.000	0.306	0.108	11.532	20.465	8.277	12.800	22.615	4.110	10.016	16.554	11.485	14.600	16.519

Table C-8: The relative abundance of selected genes in each metagenomic dataset that were specifically mentioned in the text. Values represent counts out of 1,000,000 annotations made. Gene descriptions and IDs correspond to those given in the Swiss-Prot protein database (The Uniprot Consortium).

Bin ID	POPULATION BIN BASIC INFO				104 SINGLE COPY BACTERIAL GENES				Estimated Completeness	Estimated Contamination
	n50 (bp)	Longest (bp)	Total Length (bp)	Number Contigs	0	1	2	3+		
1	87585	258416	4341484	85	5	98	1	0	94.0%	1.7%
2	44196	144885	4793253	228	21	83	0	0	89.7%	0.0%
3	167654	415574	5822639	110	4	100	0	0	95.7%	0.0%
4	36041	146424	2468837	156	14	90	0	0	81.0%	0.0%
5	20203	76412	2053920	189	12	92	0	0	83.5%	0.0%
6	54187	226195	3684989	154	2	102	0	0	97.4%	0.0%
7	142525	388009	5910590	149	12	92	0	0	89.8%	0.0%
8	65109	167976	3268372	73	3	100	1	0	95.7%	1.7%
9	5369	29935	3892756	976	69	34	1	0	54.3%	1.7%
10	12295	50046	2301754	289	18	85	1	0	81.3%	1.7%
11	51987	197629	5172412	191	2	102	0	0	98.8%	0.0%
12	58175	161485	2468333	73	80	24	0	0	37.1%	0.0%
13	10550	36899	3730785	564	27	75	2	0	66.6%	3.5%
14	11076	60265	4040674	575	39	64	1	0	65.7%	1.7%
15	4362	19644	2402663	675	58	46	0	0	53.0%	0.0%
16	16102	93823	5151448	478	57	46	1	0	68.7%	1.7%
17	6223	29378	3210355	712	52	52	0	0	48.3%	0.0%
18	12136	64006	7779547	979	44	52	8	0	72.1%	13.8%
19	9878	99463	4721202	820	45	58	1	0	65.5%	1.7%
20	13258	53656	3231835	375	31	71	2	0	83.1%	1.0%
21	6777	61204	3385891	718	41	62	1	0	55.4%	1.7%
22	27045	163169	4115219	235	30	74	0	0	74.4%	0.0%
23	20762	49785	1603225	119	49	55	0	0	44.1%	0.0%
24	218163	218163	388312	3	104	0	0	0	0.0%	0.0%
25	13861	63010	585735	55	103	1	0	0	0.9%	0.0%
26	15154	70312	961897	108	102	2	0	0	3.5%	0.0%
27	26244	147429	8218361	566	3	101	0	0	94.8%	0.0%

Table C-9: Summary statistics for each bin. Including n50 (*i.e.*, the length that >50% of the assembly is in contigs of this length or longer), the length of the longest contig, the total length of all contigs combined, and the number of contigs comprising a bin. All contigs were > 1kbp. CheckM was used to assess each bin for completedness, checking for the presence of 104 single copy bacterial genes. The number in each column represents the number of instances each of the 104 Bacterial single copy genes was found in each bin (i.e. 98 single copy genes were found once in Bin 01). Bins highlighted in green indicate those estimated to be >80% complete. Single copy marker genes that occur more than once serve as an indicator of contamination – *i.e.* that the assembly is combined with sequences from a more than one organism.

Bin ID	Length (bp)	ID Match (%)	Best Match (SILVA LSU database v119)	Database accession ID
1	2872	93.04	Verrucomicrobia;Opitutae;Opitutales;Opitutaceae;Opitutus;Opitutus terrae PB90-1	CP001032.2053384.2056251
3	568	94.37	Acidobacteria;Acidobacteriales;Acidobacteriaceae;Acidobacteriaceae bacterium KBS 83	ARMD01000022.431651.434588
6	1546	80.21	Acidobacteria;Acidobacteriales;Acidobacteriaceae;Candidatus Koribacter;Candidatus Koribacter versatilis Ellin345	CP000360.5256362.5259327
7	358	94.69	Acidobacteria;Acidobacteriales;Acidobacteriaceae;Granulicella;Granulicella mallensis MP5ACTX8	CP003130.693578.696525
8	3001	93.37	Actinobacteria;Thermoleophilia;Solirubrobacterales;Conexibacteraceae;Conexibacter;Conexibacter woesei DSM 14684	CP001854.3761310.3764306
10	996	84.74	Proteobacteria;Gammaproteobacteria;Legionellales;Coxiellaceae;Coxiella;Pseudomonas moraviensis R28-S	AYMZ01000007.155400.157795
12	487	94.25	Acidobacteria;Acidobacteria;Acidobacteriales;Acidobacteriaceae;Acidobacteriaceae bacterium KBS 83	ARMD01000022.431651.434588
27	3022	88.05	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Polyangiaceae;Sorangium;Sorangium cellulosum So ce56	AM746676.3488030.3491045

Bin ID	Length (bp)	ID Match (%)	Best Match (SILVA SSU database v119)	Database accession ID
1	1484	98.99	Verrucomicrobia;Opitutae;Opitutales;Opitutaceae;Opitutus	JQ311870.1.1484
2	319	99.37	Acidobacteria;Acidobacteriales;Acidobacteriaceae	GQ339162.1.1463
3	338	99.41	Acidobacteria;Subgroup 3;Unknown Family;Candidatus Solibacter	EF018719.1.1346
4	760	95.66	Chlorobi;Ignavibacteria;Ignavibacteriales	FR667819.1.1489
8	1498	99.07	Actinobacteria;Thermoleophilia;Solirubrobacterales;TM146	EU861899.1.1498
10	973	98.77	Proteobacteria;Gammaproteobacteria;Xanthomonadales	AY963465.1.1458
14	254	97.24	Actinobacteria;Actinobacteria;Frankiales;Acidothermaceae;Acidothermus	AB821057.1.1475
20	169	100	Acidobacteria;Acidobacteriales;Acidobacteriaceae (Subgroup 1)	FJ624922.1.1488
22	241	99.59	Acidobacteria;Acidobacteriales;Acidobacteriaceae (Subgroup 1)	GQ339162.1.1463
27	1557	96.34	Proteobacteria;Deltaproteobacteria;Myxococcales;Polyangiaceae;Sorangium	EU445217.1.1547

Table C-10: Information on bins containing ribosomal sequences. 16S rRNA gene matches were determined by using a BLAST search (word size = 18) on whole population bins against SILVA SSU (small subunit) ribosomal RNA sequences (v119 release), and 23S rDNA matches were determined by using a blast search (word size = 18) on whole population bins against SILVA LSU (large subunit) ribosomal RNA sequences (v119 release). Only matches that were greater than 100bp in length are displayed.

APPENDIX D: SUPPLEMENTAL MATERIAL FOR CHAPTER 5

D.1. Suppemental tables and figures

Sample metagenome	# Raw sequences	# Post-QC sequences	% Reads retained	Raw data size (Gbp)	Data size post-QC (Gbp)
Control 1 (15-25cm)	16921626	14009649	82.8%	5.08	3.02
Control 2 (15-25cm)	27732641	22464651	81.0%	8.32	5.56
Control 3 (15-25cm)	23357438	18245350	78.1%	7.01	4.79
Control 4 (15-25cm)	28854107	20482564	71.0%	8.66	5.65
Control 5 (15-25cm)	37835617	30044541	79.4%	11.35	6.35
Control 6 (15-25cm)	21182813	15933022	75.2%	6.35	4.26
Warming 1 (15-25cm)	24473078	19649679	80.3%	7.34	4.94
Warming 2 (15-25cm)	29401318	24324617	82.7%	8.82	5.99
Warming 3 (15-25cm)	31374460	22870499	72.9%	9.41	6.08
Warming 4 (15-25cm)	20149558	15834600	78.6%	6.04	4.07
Warming 5 (15-25cm)	23412481	18608517	79.5%	7.02	4.56
Warming 6 (15-25cm)	38005737	31682650	83.4%	11.40	6.93
Control 1 (45-55cm)	14057671	11455841	81.5%	4.22	2.88
Control 2 (45-55cm)	20247622	16323195	80.6%	6.07	3.58
Control 3 (45-55cm)	18636844	15863933	85.1%	5.59	3.78
Control 4 (45-55cm)	18322419	15359809	83.8%	5.50	3.56
Control 5 (45-55cm)	15207692	12794505	84.1%	4.56	3.22
Control 6 (45-55cm)	19174175	16153340	84.2%	5.75	3.73
Warming 1 (45-55cm)	24867267	20116318	80.9%	7.46	4.09
Warming 2 (45-55cm)	26240894	20213517	77.0%	7.87	3.81
Warming 3 (45-55cm)	22183927	18370879	82.8%	6.66	4.00
Warming 4 (45-55cm)	19747772	16796376	85.1%	5.92	3.95
Warming 5 (45-55cm)	12737442	10115066	79.4%	3.82	2.48
Warming 6 (45-55cm)	18580029	15219345	81.9%	5.57	3.36

Table D-1: Information pertaining to the sequence quality and percent data retention for all twenty-four soil community metagenomes from the 5-year CiPEHR sample collection.

Experimental plot	Soil depth	Treatment	Est. avg genome size
Fence 1	15-25 cm	Control	6929393
Fence 2	15-25 cm	Control	6699766
Fence 3	15-25 cm	Control	6345230
Fence 4	15-25 cm	Control	6288559
Fence 5	15-25 cm	Control	6130808
Fence 6	15-25 cm	Control	7500976
Fence 1	15-25 cm	Warming	6806590
Fence 2	15-25 cm	Warming	6327260
Fence 3	15-25 cm	Warming	7035674
Fence 4	15-25 cm	Warming	7438793
Fence 5	15-25 cm	Warming	7009593
Fence 6	15-25 cm	Warming	5873189
Fence 1	45-55 cm	Control	4794725
Fence 2	45-55 cm	Control	4785026
Fence 3	45-55 cm	Control	5129050
Fence 4	45-55 cm	Control	5186224
Fence 5	45-55 cm	Control	5014540
Fence 6	45-55 cm	Control	4131485
Fence 1	45-55 cm	Warming	4248245
Fence 2	45-55 cm	Warming	4162469
Fence 3	45-55 cm	Warming	4878232
Fence 4	45-55 cm	Warming	4282672
Fence 5	45-55 cm	Warming	4954275
Fence 6	45-55 cm	Warming	4799293

Table D-2: Estimated average genome size of each sample community metagenome as determined with Microbe Census. CheckM is a tool that estimates average genome size by normalizing the number of reads matching to a set of conserved single copy marker genes by the total number of reads used in query.

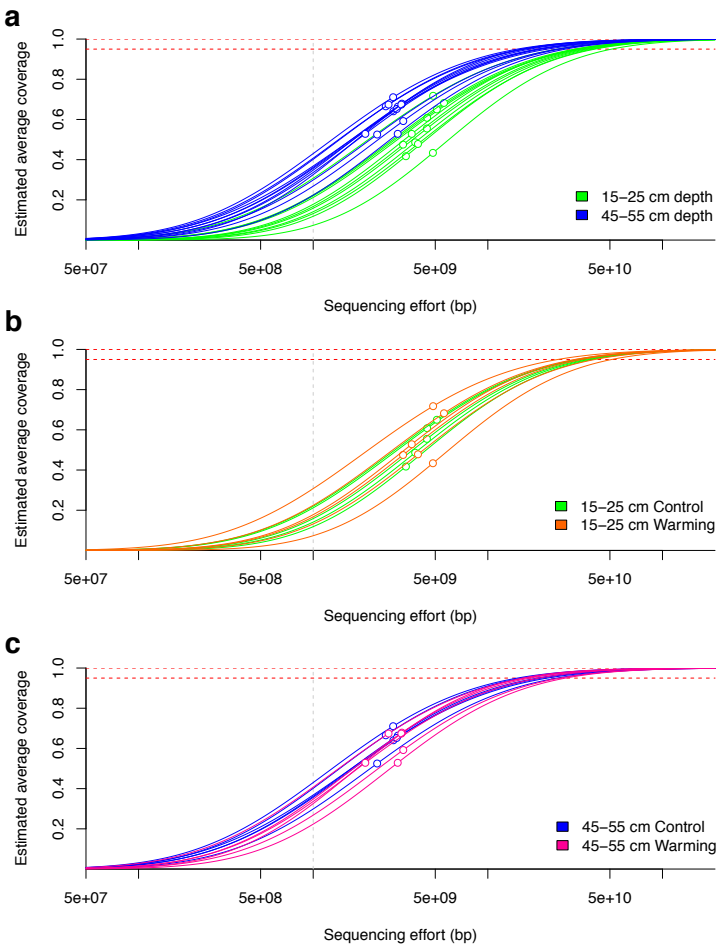


Figure D-1: Curves representing soil microbial community complexity as determined by Nonpareil. Nonpareil is a statistical tool that uses read redundancy to estimate metagenome dataset complexity and the amount of sequencing effort needed to achieve a desired level of coverage (Rodriguez-R *et al.,* 2018). Different plots are shown to compare (a) all 5-year sample metagenomes, (b) 15-25 cm sample metagenomes only, and (c) 45-55 cm sample metagenomes only. Circles on curves represent the coverage of the actual sequencing depth for each dataset in relation to the entire curve (projection for complete coverage after the circle). Curves positioned on the right represent more sequence diverse metagenomes than curves positioned on the left.

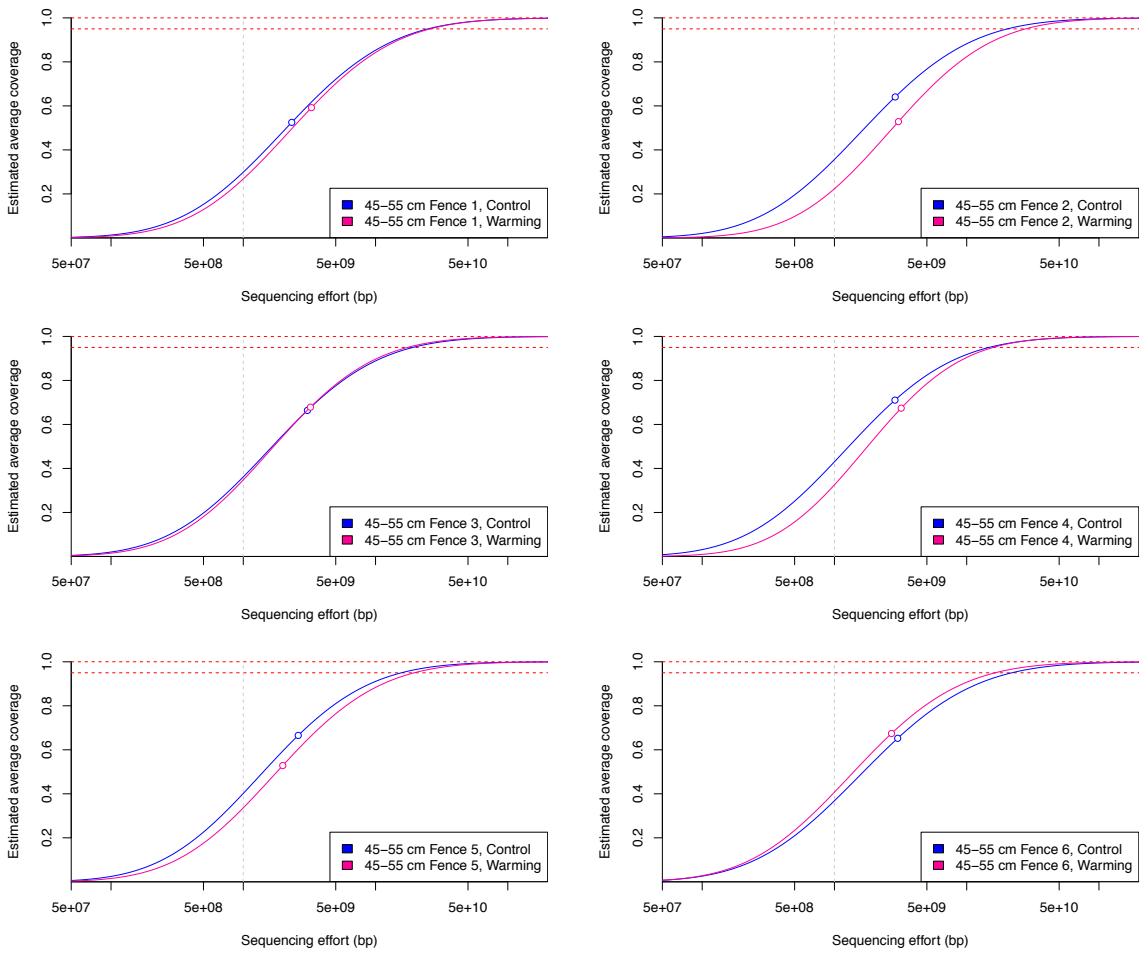


Figure D-2: Curves representing pairwise comparisons of warmed and unwarmed soil communities after 5 years of experimentation at at 45-55 cm. Each plot shows the differences between warmed and unwarmed samples corresponding to the six different fences (i.e., experimental plots). See Figure D-1 legend for additional details.

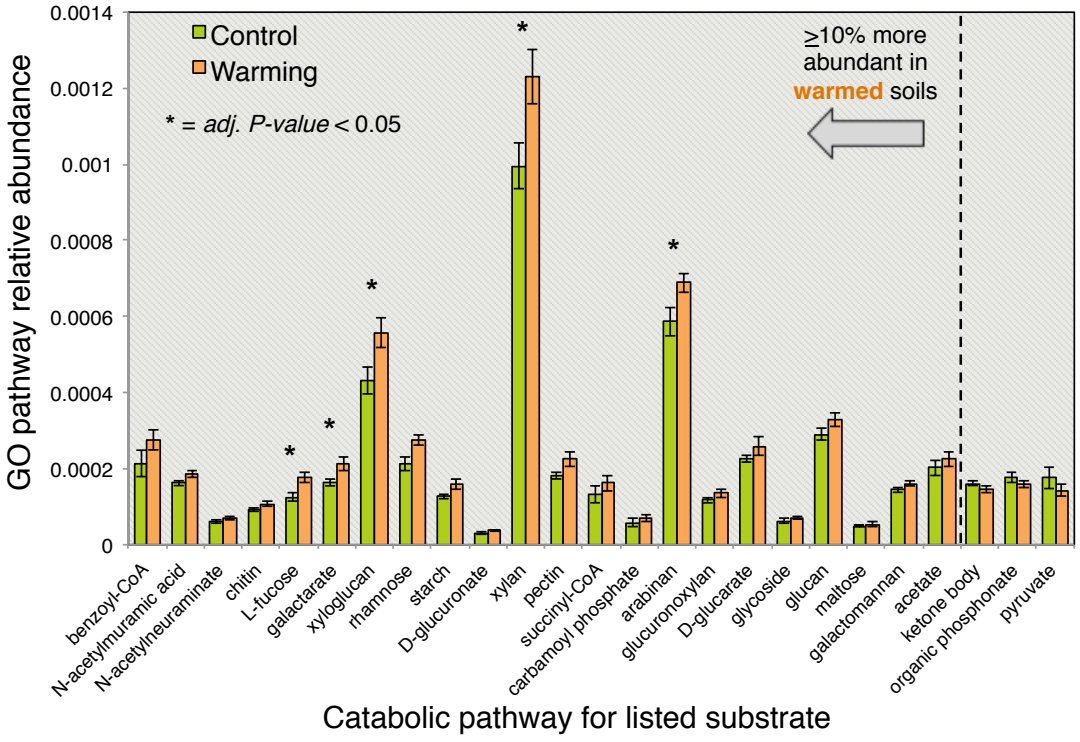


Figure D-3: Shifts in the relative abundances of genes involved in the catabolism of various organic matter substrates at 15-25 cm. The relative abundancess of catabolism pathways was determined by matching Swiss-Prot references to Gene Ontology (GO) Biological Process terms, and dividing the number of annotations to each pathway by the total number of annotations to Swiss-Prot. Bars reflect the mean relative abundance for each sample group (unwarmed, warmed) at 15-25 cm (n = 6). Bars represent the standard error of the mean. Only pathways that differed between warmed and unwarmed soils by $\geq$ 10%, on average, are shown.

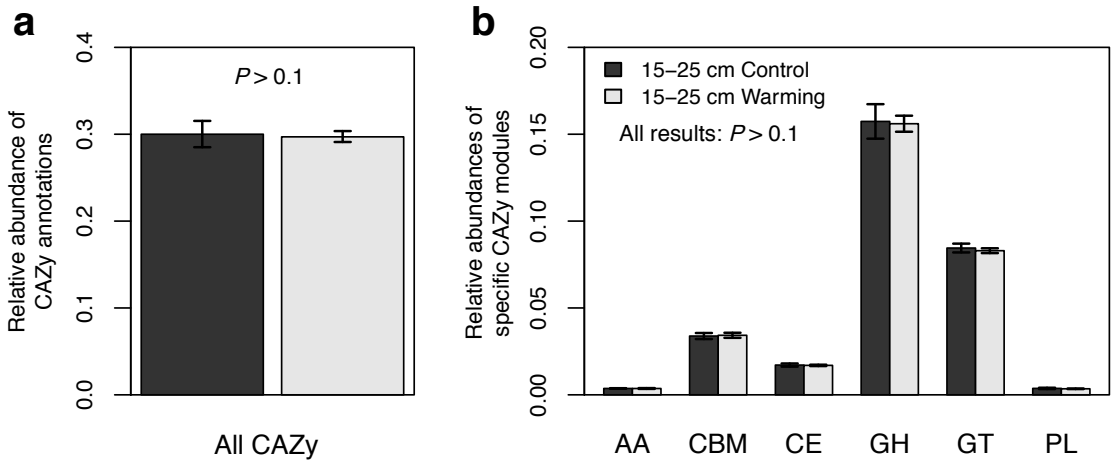


Figure D-4: Shift in carbohydrate utilization (CAZy) genes as an effect of of experimental warming after 1.5 years. Relative abundances of (a) all CAZy annotations, and (b) CAZy modules: Auxiliary Activities (AA), Carbohydrate Binding (CBM), Carbohydrate Esterases (CE), Glycoside Hydrolases (GH), GlycosylTransferases (GT), and Polysaccharide Lyases (PL). Underlying values represent the mean relative abundance of each category for warmed vs. unwarmed soils.

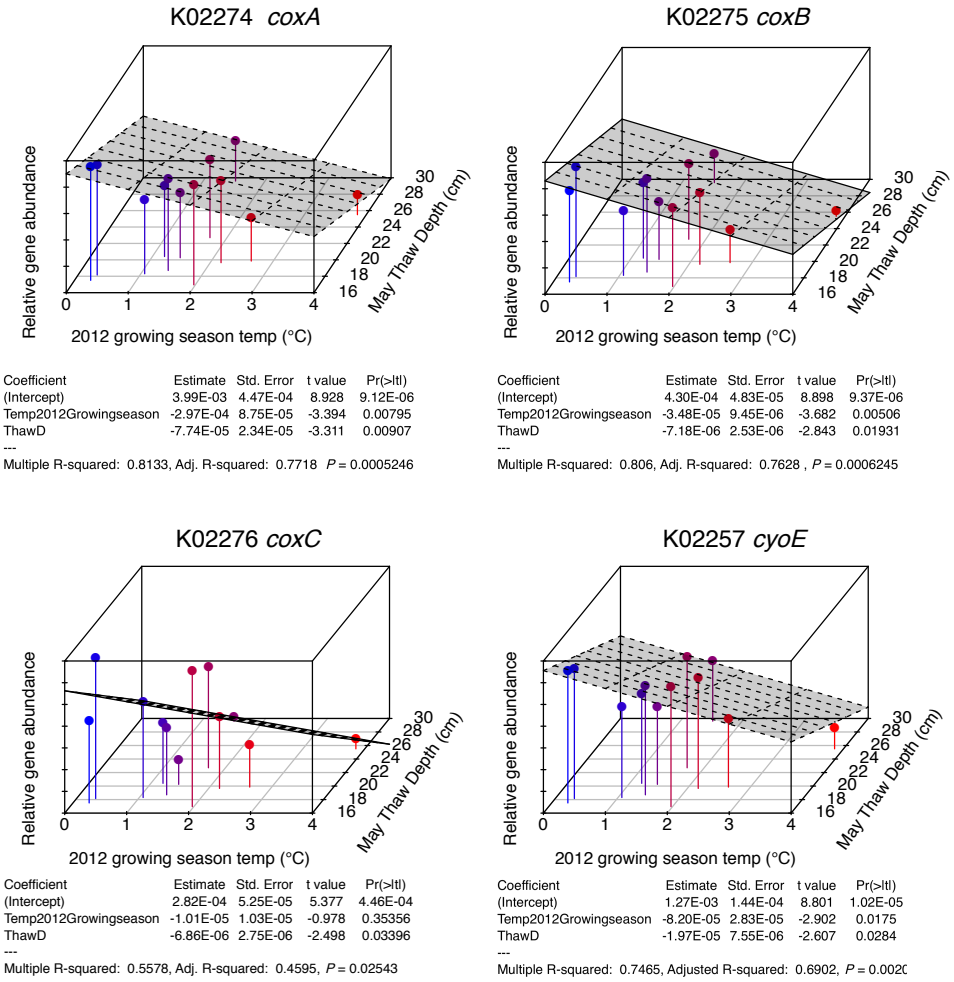


Figure D-5: Idependent and additive effects of thaw depth and average summer temperature on the relative abundances of cytochrome c oxidase genes in 45-55 cm soils. Three-dimentional plots were made using the R package 'Scatterplot3d' and illustrate relationships between average summer soil temperatures, thaw depth at time of sampling, and the abundance of cytochrome c oxidase genes (title of each plot reflects the KEGG Ontology identifier, as well as the common gene identifier). R function lm() was used to fit multiple variables to determine assess their independent and combinatorial effects. Individual data points are colored by average summer temperature, which proceed from blue to red to reflect low vs. comparatively higher soil temperatures

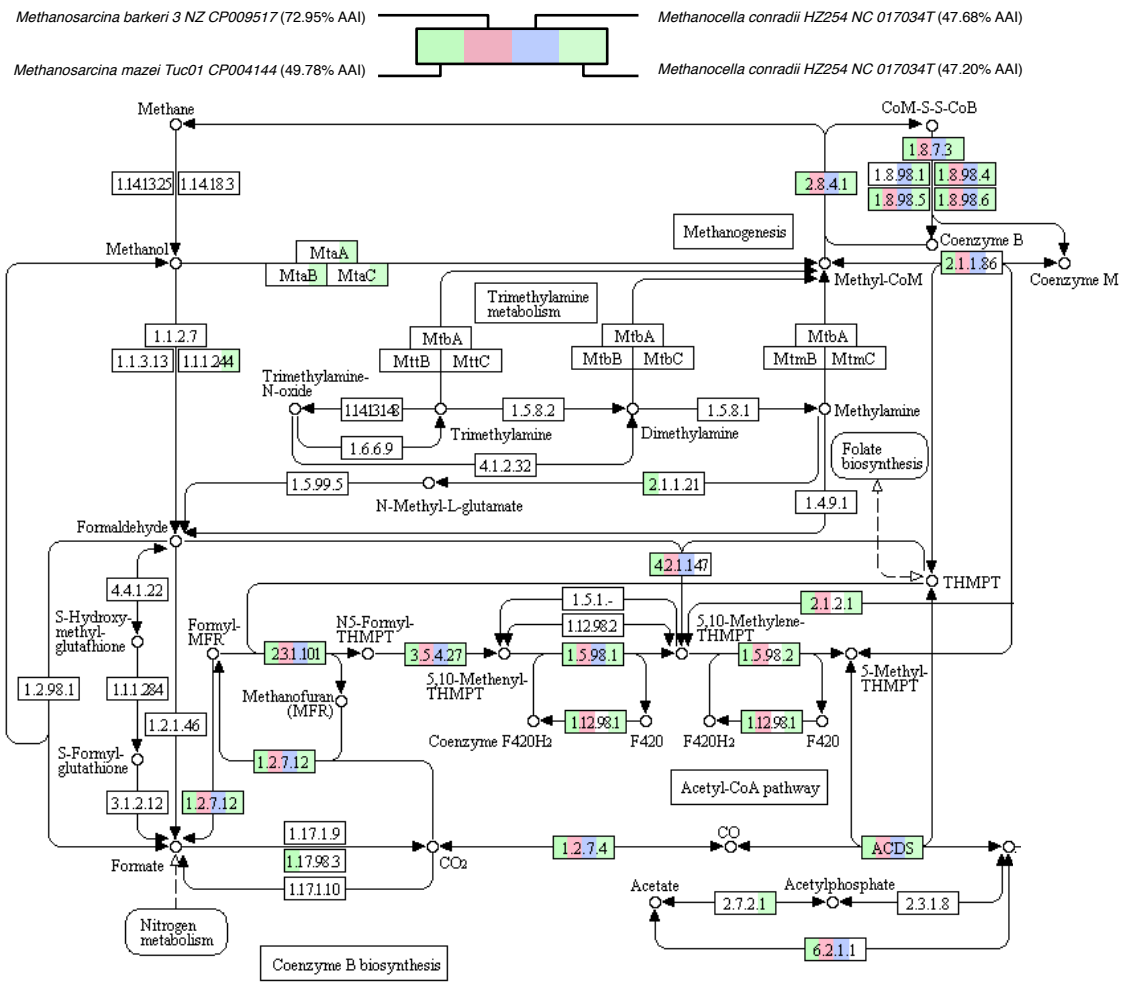


Figure D-6: Taxonomic assignment and methane generating functions of recovered archaeal MAGs. Diagram represents the KEGG reference pathway 'Methane metabolism' (map00680). Coloration of boxes is used to denote the presence of each function in each of the four non-redundant methanogenic MAGs (see top key). MAG taxonomic assignment was determined by using The Microbial Genomes Atlas (MiGA) webserver (http://enve-omics.ce.gatech.edu:3000) (Rodriguez-R et al., 2018). The percentage accompanying each taxonomic classification reflects the highest AAI (average amino acid identity) with a collection of known genome references.

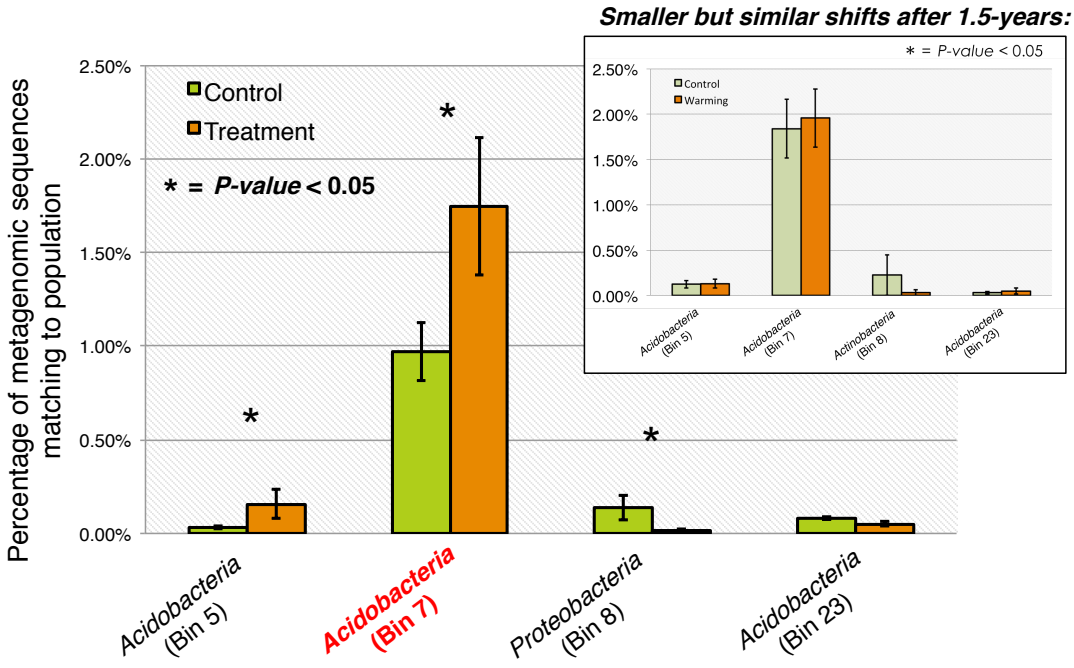


Figure D-7: Shifts in the relative abundances of dominant MAGs at 15-25 cm after 1.5 and 5 years of experimental warming. MAG relative abundances were determined by dividing the number of short-reads matching each MAG at > 80 bp and > 98% nucleotide identity by the total number of reads used as query. Bars reflect the mean relative abundance for each sample group (unwarmed, warmed) at 15-25 cm (*n* = 6). Bars represent the standard error of the mean.